

PT Novel methods and kits for detecting the presence of *Bacillus anthracis*
 PT In a test sample -
 XX
 PS Claim 2: Page 59-60; 62pp; English.
 XX
 CC The present sequence represents a bacterial surface array protein (SAP).
 CC SAP is used in the method of the invention. The specification describes
 CC a method for detecting the presence of *Bacillus anthracis* in a test
 CC sample. The method comprises contacting the sample with a capture
 CC reagent and detecting whether the a SAP is bound to the capture
 CC reagent, which is indicative of the presence of *B. anthracis* in the
 CC sample. The method is useful for detecting the presence or absence of
 CC *B. anthracis* in a test sample.
 XX
 SQ Sequence 785 AA:

Query Match 100.0%; Score 3885; DB 22; Length 785;
 Best Local Similarity 100.0%; Pred. No. 2.1e-203;
 Matches 785; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGKTPDVADHMGIDISINYLVEKAVGKNDKGFEPGKELTRAAATMAOITLWPIDK 60
 DB 1 agktfpdvadhwgidsinylvekvagvndkgmfepgkeltreaaatmmaqilnlpidk 60
 QY 61 DAKPSFADSGOWYTPETIAAVEKAGVINGTGNGFEPNGKIDRVSMASLVEAYKLDTRKVN 120
 DB 61 dakpsfadsqgwytpetiaavekagvikgtngfepngkidrvsmasllveaykldtrkvn 120
 QY 121 GTPATKFKDLETLMNGKREKANTVELGTSVGTGDQWEPKRYTKAAAOFTAKTDKQFT 180
 DB 121 gtpatkfkldletlmngkreaanlvelgtsvgtdqwepkrtvkaaaoflaktkqft 180
 QY 181 EAAKVESAKAVTTOKVEVFKSAVEKELTKEDIKVYNKANNDKVLVEYTLSEDKRSATVE 240
 DB 181 eaakvesakavttgvevfkavekklkedi kvnknndkvlyveytlsedkrsatve 240
 QY 241 LYSNLAAKOTYTVDNKKYKTEVANGSLAEAKTIEKADOTVVADEPTALQFTVKDENGTEV 300
 DB 241 lysnlaakqlycvdnkkyktevavngslaeaktiemadclvadeptalqftvkdengelev 300
 QY 301 VSPGIEFVTPAAERINAKGETITLAKGTSTYVKAAYKKDKGRVVAESKPKVKSAGCAVVAS 360
 DB 301 vspegiefvtpaaeklnakgetitlakgststvykaykkdkgrvvaseskpksaagcaavas 360
 QY 361 ISNMTPVAEQNKADFTSKDFKONKNVYEGDNAYVOVELKDQFNNAVTTGKVEYESLNTVEAY 420
 DB 361 isnmtpvaeqnkdftskdfkgnknvyegdnayvvelkdqfnavttgkveyeslntevay 420
 QY 421 VDKATGKVTYLSAGKAPYKVTVKDSKALVSHVETEAFAOKMKIKIEKTVNALSTK 480
 DB 421 vdkatgkvtvlsagkapykvtvkdsakalvshvetleaafakmkikiektvnaistk 480
 QY 481 DYTDLKAKKAPVLDQYGEFTAPVTVKVLDRDGKELKBOKEAKYVNNRELVNNAAGQAGN 540
 DB 481 dytdlkakkapvldygeftapvtvkvldrdgkelkbokeakyvnnrelvlnnaagqagn 540
 QY 541 YTYVLTAKKSEBEKATLAEKAPGAFSEFVRCGLDELDKVTVEBNQNMAMTVSLPV 600
 DB 541 ytyvltaksebeekatlalekapgafsefvrcldeildkvtvebnqnmamtvslpv 600
 QY 601 DANGLVKGAEAELKATTNNKEGKEVDATDAOVTONNSVITVGGAKGERTYKTVVL 660
 DB 601 danglvkgaeealkattnnkegkevdatdaovtonnsvitvggakgertykvtvvl 660
 QY 661 DGLITTHSFKVVDTAPTAAGLAVEFTSTSLKEVAPNADLKAALMLILSDVGPATTTAKA 720
 DB 661 dglltthsfkvvdtpataglaveftstslkevapnadlkaallilsvdgpatttaka 720
 QY 721 TAANVFVSADTVNVAENGNGVAKGATSTYVKULTVYKKGKSGKVEFEDKAVOVAVSIRKA 780
 DB 721 taanvfvsadtlvnvaengngvkgatstiyvknltvvkkgksgkvefedkavovasirkea 780

QY 781 KPATK 785
 DB 781 kpatk 785

RESULT 2
 ID AAW22863 standard; Protein: 921 AA.
 AC AAW22863;
 DT 27-FEB-1998 (first entry)

DE *Bacillus stearothermophilus* s-layer protein sbs-B.

KW 5-layer; sbs-B; Vaccine; adjuvant; carrier; hybridisation assay;
 KW molecular spinning nozzle; molecular laser.

OS *Bacillus stearothermophilus*.

Key Location/Qualifiers
 FT Peptide 1..31
 FT /label= sig-peptide
 FT Peptide 32..921
 FT /label= mat-peptide

DE19603649-A1.

07-AUG-1997.

01-FEB-1996; 96DE-1003649.

01-FEB-1996; 96DE-1003649.

(LUBI/) LUBITZ W.
 (SLEY/) SLEYTR U.

Kuen B, Lubitz W, Sleytr U;

WPI: 1997-394558/37.

DR N-PSDB; AAT75488.

PT Preparation of S-layer proteins by expressing sbs-A gene in Gram
 PT negative bacterium - or new sbs-B gene in any host, also new
 PT recombinant proteins containing heterologous inserts, e.g.
 PT epitope(s), useful as vaccines and adjuvants

PS Claim 26; Pages 19-23; 31pp; German.

XX
 CC The present sequence is the *Bacillus stearothermophilus* PV72
 CC S-layer protein, sbs-B. S-layer structures can be used as vaccines
 CC or adjuvants, particularly when they include a bacterial ghost that
 CC may contain additional epitopes in its membrane. Other uses of
 CC recombinant sbs-B, depending on the nature of the inserted peptide,
 CC are as an universal carrier for biotinylated reactants for use in
 CC immunological or hybridisation assays (the insert is streptavidin),
 CC to induce immune responses (epitopes), as a reagent for removing
 CC cytokine or toxin from serum (antigenic epitopes) and as a molecular
 CC spinning nozzle (polyhydroxybutyrate synthase) and as a molecular
 CC laser (luciferase).

XX Sequence 921 AA:

Query Match 10.8%; Score 419.5; DB 18; Length 921;
 Best Local Similarity 24.0%; Pred. No. 4.8e-15;
 Matches 211; Conservative 133; Mismatches 332; Indels 203; Gaps 40;

QY 4 TFPDVPADHMGIDISINYLVEKAVGKNDKGFEPGKELTRAAATMAOITLWPIDKAK 63
 DB 33 sftdpvady--kdaidflvsqatkgctekfygydeltridaavilarvikldvnaekd 90
 QY 64 PSFADSGOWYTPETIAAVEKAGVINGTGNG--FEPNGKIDRVSMASLVEAYKLD----- 116

```
Db      91 agfdtrpdkd-rakyyvalveagvlnngkapgkfgaydpdlrtvemaklianykkladadvkl 149
      117 --TKVNGTPATKFEKDELTLMNGKEKANIIVELGISVGTGDMPEPKTVTTRAAAOFTAKT 174
      150 pfldvndtwapykal----ykvevtrklkhqgasvnt-----khltrfdqfyra 198
Qy      175 DKQGTBAAVESAKAVTTQKVEYKESKAVEKLTKEDIKVTNNKANNKVLKVEVTSDEK 234
      199 vni--navpeivevtaavstvtkvfntqj----advdfntfaidnglvtvkalisrdk 251
Qy      235 RSATVELYSMLAKQOTVTVNKGK--KTEVA-----VGSIL-EAKTIMADOTVVADE 284
      252 ksvevvvnhkftmngeytltatgiknlkgetakeltgkfwsvgdavtvalnmsllyge 311
Qy      285 PTLAQFTVKDENGTEVY-----SPEGIEFVTPAAEKINAKGEITLAKGTSTTVKAVY 336
      312 esgl--lvkqgdqgdvvgakvveltsntlnlvvssgevsaaavtvalvkpgtdavtakvt 369
Qy      337 KKDGKVAVES-----KEVKVSABGAVAASISNMVTAEO-----NKADFTSKDFKQNNK-- 384
      370 lpdgvvltntfkkvltvepvvgvngqgftlvdnlsnapntvafnkkaekvlsmfagelctv 429
Qy      385 -VYE---GDNAVVOVELKDOENAVTTGKVEESLNTFVAVVDKATGVTVLS-----AGKA 436
      430 amydtkngdpdectpvdtkd-----atvrslnplataaingselivtanagsgska 480
Qy      437 PVKTVVNDKSGKALVSHVTEIEFAQAKMKDIKEKTNNVALS-----TKDVTDLKVK 488
      481 sfetrlkdn-----tkrtfvdvkkdpvlqldikvdatssvklstdeavvggvegvngtklik 535
Qy      489 APVLDOYGEKEF-----TABVTVKVLDDK-----GKLEKEQ----- 518
      536 vsavdgvgkeltkfgtkgvtrtnteglviknvsndhtlfdsgnsatdgvvavatkdkl 595
Qy      519 ---KLEAKYNNRELVLNAAOGAENVTVVLAKSGEKEAKAT-LALELKAPGASKEVEVR 574
      596 vngvngvkkfkn-----asdtptscctltlvnnvknkadtpvgldvdp--skldvn 646
Qy      575 GLDPELKYVTEENOKNAMTVSLPVDPANGLVLKGAAEALIKVTT-----NKEGKEVD 628
      647 apnt---astadvdfinfesvelytldnsng-----rrqkvvptatcltvgtkkkkkn 696
Qy      629 ATPDAQTVQNNSVITVVGAGAKAGETVTVTVLGDKLITTHSEKVVVDTPAPAKGLAVEFTS 688
      697 gnvly--fkgngeeltstssstgn-----vdg-----taegnt----- 727
Qy      689 TSUKEVAPNADIKALNLTLSVDCVPATTAKTAS-NVEFVSADTNVVAENGTVGAKGAT 747
      728 ---krl-pgkylinsa-----svpasatvatpvtvklinsdndltfeelifgvldpt 775
Db      748 SIYVKN-----LTVNRDKGEQKVEFPKAVOVAVSIEKA 780
      776 ql-vkdedinefiavskkaekndgylhnpkl---vtvkda 810
Qy      776 ql-vkdedinefiavskkaekndgylhnpkl---vtvkda 810
RESULT 3
ID AAB10626 standard; Protein; 921 AA.
XX AAB10626;
DT 08-JAN-2001 (first entry)
XX B. stearothermophilus sbsb protein.
XX S-layer gene: sbsb; antibacterial; vaccine; adjuvant; bioreactor;
XX poly(hydroxyalkanoate) synthesis.
XX Bacillus stearothermophilus.
XX Key Location/Qualifiers
FT Peptide 1..31
```

```
FT Peptide /label= signal_peptide
FT 32..921 /label= mature_peptide
XX DE19903345-A1.
XX 03-AUG-2000.
XX 28-JAN-1999; 99DE-1003345.
XX 28-JAN-1999; 99DE-1003345.
XX 28-JAN-1999; 99DE-1003345.
XX (LUBIT/) LUBITZ W.
XX Lubitz W.
XX WPI: 2000-533868/49.
XX N-PSDB; AAA71798.
XX Host cell, useful e.g. as bioreactor for production of
XX poly(hydroxyalkanoate), containing two or more recombinant
XX polypeptides, with at least one in carrier-bound form -
XX Disclosure; Page 20-22; 26pp; German.
XX This invention describes a novel host cell (A) comprising at least two
XX functional recombinant polypeptides (1), at least one being in carrier
XX bound form. The products of the invention have antibacterial activity.
XX (A), or, where bacterial, their ghosts (B), are useful as vaccines or
XX adjuvants (specifically for presentation of immunogenic epitopes of
XX pathogens or autologous immunostimulatory polypeptides, e.g. cytokines),
XX or preferably, as enzyme reactors for performing a cascade of reactions,
XX specifically synthesis of poly(hydroxyalkanoate). Localization of
XX individual (1), specifically enzymes, in separate cellular compartments
XX avoids adverse reactions between products and substrates, when being used
XX as bioreactors. (1) can be produced in carrier-bound form without loss of
XX function. This sequence represents the Bacillus stearothermophilus
XX S-layer protein sbsb which is used to illustrate the method of the
XX invention.
XX Sequence 921 AA:
Query Match 10.8%; Score 419.5; DB 21; Length 921;
Best Local Similarity 24.0%; Pred. No. 4,8e-15;
Matches 211; Conservative 133; Mismatches 332; Indels 203; Gaps 40;
Qy 4 TFPVPADHMGIDSINLVLEKGAVKGNDKGMFEFGKELTRAEAATMAQIQLNPIDDK 63
      33 sftlvapgy--kdaidlvtsgakkgkteckfygydeltridaavliaryklidvndakd 90
Db 64 PSFADSGOWTTPPIAVERAGVIRKGTGNG-FEPNGKIDRVSMASLLVEAYKID----- 116
      91 agfdtrpdkd-rakyyvalveagvlnngkapgkfgaydpdlrtvemaklianykkladadvkl 149
Qy 117 --TKVNGTPATKFEKDELTLMNGKEKANIIVELGISVGTGDMPEPKTVTTRAAAOFTAKT 174
      150 pfldvndtwapykal----ykvevtrklkhqgasvnt-----khltrfdqfyra 198
Qy 175 DKQGTBAAVESAKAVTTQKVEYKESKAVEKLTKEDIKVTNNKANNKVLKVEVTSDEK 234
      199 vni--navpeivevtaavstvtkvfntqj----advdfntfaidnglvtvkalisrdk 251
Qy 235 RSATVELYSMLAKQOTVTVNKGK--KTEVA-----VGSIL-EAKTIMADOTVVADE 284
      252 ksvevvvnhkftmngeytltatgiknlkgetakeltgkfwsvgdavtvalnmsllyge 311
Qy 285 PTLAQFTVKDENGTEVY-----SPEGIEFVTPAAEKINAKGEITLAKGTSTTVKAVY 336
      312 esgl--lvkqgdqgdvvgakvveltsntlnlvvssgevsaaavtvalvkpgtdavtakvt 369
Db 337 KKDGKVAVES-----KEVKVSABGAVAASISNMVTAEO-----NKADFTSKDFKQNNK-- 384
```



```
Db 480 sfefvfkdn-----tkrtfvdvkkdpvlgdtkvdatstvklisdeavvggvegvnqklix 534
QY 489 APVLDOYKEF-----TAPVTYVLDKD-----GKELKEQ----- 518
Db 535 vsadqgqkklfktgkvvtltnteglviknvsndltidfsqnsatcdqfvvaatkxi 594
QY 519 --KLEAKYVRELVLNAGOGAGNYTVLTAKEGSEKAKAT-LALELKAPGAFSFEVR 574
Db 595 vngvvevfyfn-----asdtlptstklitvnnvvnkakatpvgldivap-----seidvn 645
QY 575 GLDDELKYYVEENKKNMTSVLPVDANGVLKGAELKAYTTTNGKEGVADTDQY 634
Db 646 aptn---astadvdfinfevseivlylidsnqrk-----kvpt---attlvgtndyv 692
QY 635 TVQNNSVTVVOCG-----AKAGETVYVTVVLDOGLITTH-SFKVVDTPATKGLAVEFTST 689
Db 693 evnngvniqfkgndeltlittssstvnvdtadg--itkrlipkylnsavspsatvatspv 750
QY 690 SLKEVAPNADIKALNLILSYDGVPAATTAKATASNVFEVSADFNVAENGTVGAKGATSI 749
Db 751 tvklinsndltfeellfgvid--ptglvkdedln-eflav-----skaakndgyl 798
QY 750 YVKNLTIVYKD 759
Db 799 ynkplvtvkd 808

RESULT 5
AAR80530
ID AAR80530 standard; Protein: 1252 AA.
XX
AC AAR80530;
XX
DT 22-DEC-1995 (first entry)
XX
DE B. sphaericus SLP.
XX
KM Surface layer protein; SLP; fusion protein; vaccine; antigen;
XX
OS Bacillus sphaericus.
XX
FH Key Location/Qualifiers
FT Peptide 1..30
FT /label= sig_peptide
XX
PN MO9519371-A2.
XX
PD 20-JUL-1995.
XX
PF 13-JAN-1995; 95WO-EP00147.
XX
PR 14-JAN-1994; 94GB-0000650.
XX
PA (SOLV ) SOLVAY SA.
XX
PI Deblaere RY, Desomer J, Dhese P;
XX
DR WPI: 1995-263827/34.
XX
DR N-PSDB: MAQ99430.
XX
PT Host cell expressing surface layer protein fusion protein - used for
PT host presentation of antigens and vaccine prodn.
XX
PS Disclosure; Fig.6; 95pp; English.
XX
CC A probe based on the N-terminal sequence of B. sphaericus P-1 (LMG
CC P-13855) surface layer protein was used to screen an HindIII-
CC generated library to isolate the slp gene. Promoter regions
CC of the gene are used in genetic constructs providing surface
CC expression of heterologous proteins in P-1 hosts.
XX
```

```
SQ Sequence 1252 AA:
Query Match
Best Local Similarity 22.5%; Score 372.5; DB 16; Length 1252;
Matches 227; Conservative 129; Mismatches 373; Indels 281; Gaps 46;

QY 16 DSNILYVEKGAVGKNDKMPFGKELTPRAEAATMAQILNLPIDKAKPSPAD-SOGQWY 74
Db 45 eavgalvqgyigdtngnfnplntvtragaelftkaleangdv--dfkdvkagawy 102
QY 75 TPPTAAVEKACVING-TGNGEFPNGKIDRVSMASILVEAYVLDTKVNTPATRKOLETL 133
Db 103 ynsilaavanglfigvsatefapnksilrseaakllveafiglegad--lsetadassqv 159
QY 134 N-WGKEKANILVEIGISVGT-GDQWEPKRTVTKEAQFIKT-DKQGTAAKVESAKA 190
Db 160 kpwaakkyieivanglfigtdanklnpnmslitrqdfalvfrtrvdkegeqpeeaafvka 219
QY 191 VTTQKVEYKFAVEKLTK-----EDIKVTN--KANNDKVLV-----KEVTISED 233
Db 220 inntvevtfeevtnvgalnfklegleiknasvkqtnkvvltteaqfdekeyvltld 279
QY 234 KRSATVELIYNLAQKQTYTVVNNKVGKEVAV-GSLEAKTITMADQYVAADEPTA---LQ 289
Db 280 --geligfigkvga--vvpktvelvssavqklggevkvqakvtvaeqgskagjprt 332
QY 290 FTVKDENGTEVVS-----PEGIEFTVPAA----- 313
Db 333 ftvpgnmndgvvprltlgealnneegiatyslyrykegidevtayatactdrskfsjgyvfwg 392
QY 314 -EKINAKGEIT---LAKGSTTVYKAVYK-KDGKV-----VAESKEVYV 351
Db 393 vdlilsveevltgassvnnnganktykvtknpkctgpeakctfnvgfvenmmvntadkvana 452
QY 352 SAEGAAVASISNMWYVAEGKNADFTSKDEKNNKYEGONAVVOYELKQ-----Q 400
Db 453 tvngvkalqisngtalid--aaqiltidskgeatftvsgtnaavtppvylstnnstnkk 510
QY 401 FNA-----VTTGKVEYESINTEVAV---VDKATGKVTVLSAGKA-PVVFYVKD----- 444
Db 511 ysaasalqctaskvrtfaelvaeytleltradne-gevaalgtntngreykvvkkdavnlak 569
QY 445 -----SKGKALVSHITVEIEAF-----AQAKAKDIKLENTN----- 474
Db 570 neivnvaftedkdrvstvnakfyvtdqtdcavayftgdka-kqlsv-ktbdkgeatfviq 627
QY 475 -----VALSTKDYVTDLAKKAPVLDQYKEFTAPVR-VKVLDDKDGKELKEOKLEAKYV 525
Db 628 sdtvndyapciawidintsdakgqldiegepkavapisyfqaayldgsaalkayk---ksd 684
QY 526 NRELVLNAGOGAGNYTVLTAKEGSEKAKATLALAKPAGAFSFEVRSJDTDELKYYT 585
Db 685 lmkavtkfdgsetavfaaelvngsgkvvtgtsl-----kkatyctlyntgndikvndqvl 739
QY 586 EENOKNMTVSVLPVDANGVLKGAELAE-----LKVYTTT---KEGKEVADTD 631
Db 740 spnrsyvtlyea-tlssctglvtlpaknlevtsvqdkttavvlatgvlavhpdgdyafaa 798
QY 632 AQQTVY-----QNNSVIIVGO-----GAKAGETVK----- 655
Db 799 keatatfatnevpnsygtvagtqfntadsqnsnslwfaqknpykayagvsqkykyfygan 858
QY 656 -----VTVVLDOGLITTHSPFVVVDTAPAKGLAVEFTSTS 690
Db 859 gnevrfgeaawaalltqyategkvltisynvgdvtc--fkvis-----avnsstea 907
QY 691 LKEVAPN---ADLKAALNLISVDGVPATTAKATASNVFEVSADTNVYA---ENGTVGAK 744
Db 908 ikpvaoptpaaprtfgaltltpaagglvltlcatnltgsladadlnsatltvdcatsvik 967
QY 745 GATSIYVKNLTIVLVGDKGKQKVEFPDKAVQ-----VAVSIKEAKPA 783
```

Db 968 dsannsl-bltivetganyv-fattvgaqclsltagcltvtlyadakna 1015

RESULT 6
AAR77673
ID AAR77673 standard; Protein; 1228 AA.
XX
AC AAR77673:
XX
DT 16-JUL-1996 (first entry)
XX
DE S-layer protein encoded by sbas gene.
XX
KW sbas gene; S-layer; cell surface layer; expression.
XX
OS Bacillus stearothermophilus.
XX
FH Key Location/Qualifiers
FT Peptide 1..30
FT /label= sig_peptide
FT Protein 31..1228
FT /label= mat_protein
XX
PN DE4425527-A1.
XX
PD 25-JAN-1996.
XX
PE 19-JUL-1994; 94DE-4425527.
XX
PR 19-JUL-1994; 94DE-4425527.
XX
PA (VOGE-) VOCLEBUSCH GMBH.
XX
PI Lubltz W;
XX
DR WPI; 1996-077933/09.
DR N-PSDB; AAT08695.
XX
XX
PT Nucleic acid encoding signal peptide of Bacillus stearothermophilus
PT S-layer protein - which has a lysine content of at least 10 per
PT cent.
XX
XX
PS Disclousure; Page 6-7; 12pp; German.
XX
CC The claimed signal peptide encoding sequence (see tag e of AAT08695)
CC is pref. operably linked at the 3' end to a protein encoding
CC sequence. The protein is pref. the S-layer protein (see tag f of
CC AAT08695). At the 5' end, the signal peptide encoding sequence is
CC pref. linked to an expression control sequence, pref. the sequence
CC given in AAT08696.
XX
SQ Sequence 1228 AA;

Query Match 6.4%; Score 248.5; DB 17; Length 1228;
Best Local Similarity 22.4%; Pred. No. 1.3e-05;
Matches 201; Conservative 109; Mismatches 309; Indels 279; Gaps 44;

OY 55 NLPTRKDAKPSRDSQOGWYTPFIAVEKAGYIKGTGNGFEPNGKIDRVSNASLIVEAYK 114
DB 332 dvlslstdgkltlvdas---tpfennteykvvkgik---dkngkefkedaftfkrlinda 375
OY 115 LDTKVNCTPAT-----KFKLETLIN-----WGK----- 137
DB 376 vltqyfgtncvntsvnaagltfdtdcltvtvfdkllepelvsnsvltldvetygkripv 435
OY 138 ---EKANILVELGTSVGTGDQME---PKRTVT---KAAEAOFIAKTOKOFTEPAKVES 187
DB 436 laastgstltltlkealvtgkykialnvnktlgyneayelv-----ftanasapcv 489
OY 188 AKAYTT-----OKVEVFSKAVEKLTEDIKVT 215
DB 490 ataptlgtclstlsgltnvngklaggvneaglytpglfttftatkldestladnftvl 549

OY 216 NKANNDKVLKVEVLTSEDKRSATVELYSNIAKQRTYDVNKGKTE--VAVGSLEAKTI 273
DB 550 vekesgtvvaselkynadakmvtlvpkadlkentllykikikglsdkgielgvtnekly 609
OY 274 EMADQTVVADDEPTALQFTVKD--ENGTEVVSPEGIEFVTPAAEKINAKGRTITLAKGST-- 330
DB 610 eefktgdlta--plvlsvtskngdaglkvc--eaqetvtfseelnltfnatlvsgstllyg 665
OY 331 -----TWAAVYKKDG---KVVAESKE----- 348
DB 666 qvavvkgaganlaltasdlipasaevatgqdgtykvkvaanqlerngykllvfgkgata 725
OY 349 -VKVSAEGAANVAVSISNMWTAVEQNKADPFSKDKRONNKVYEGONAVQVELKQFANVTTG 407
DB 726 pykdaaanatlatalnyltlttegg-dvtarply--tkvylgds-----lkda-davlt- 773
OY 408 KVEYESLNTAVAVDVKATGKVTVLSGKAPVAVTV-----KDSGKALVSHTEIEAFA 461
DB 774 -----ltnvdaqq---kftlgfseelktssg-slyvgkvtevekl 809
OY 462 OKAMDDIKIEKTNVALSTVDYDVLKVKAPVLDQYKREFTAPVTVKVLDDKGRLEKQKLE 521
DB 810 nngwvd---agtgctvsv-----apktlangkvlaaavtlgldnndkdak-lrlv 856
OY 522 AKVNRRELVLNAAQO-----EAGNTYTVVLTAKSGEKEAKATLALELKAPGAFSK 570
DB 857 vdkssdtdgadvaagnvikedlllynswrhvtvasvkaadddq-----naaaatp- 908
OY 571 FEVRGLDTELD--KYTEENQKNAMTVSVLPVDANGVLVGAEAELKVTYTNKEGKV 627
DB 909 -----tstaiddtksljvefne-----ldlaevkpenlvvdaagnavagvtaldgs- 956
OY 628 DATDAQVTYQNNSVITVGOGAKAGETKYVTVLDG-----KLITTSFKVVDTPAP 677
DB 957 -----tnkfvtlpsqelkagltvsvtl--dgvrckvgnltlskylt--sfktsvanp 1003
OY 678 TAKGLAVEFTSTLKEVAPNADLKAALNLTISVDGP-ATPAKATAASWVERVSAD-TPNVV 735
DB 1004 t-----lsslsisdgavvnd-rsktltdfsdsvpnpitllkkgdstltnlyclvnm 1055
OY 736 AENGT---VGAKGAT--SIYKKNLTIVKDGK----EOKVEFDKAVOVAVSIRKEAKPA 783
DB 1056 nenklyklyvfhkyvldefgtgylavaskdfgtgtdidskvtf---lgsvatdevkpa 1110

RESULT 7
AAM22862
ID AAM22862 standard; Protein; 1228 AA.
XX
XX
AC AAM22862;
XX
DT 27-FEB-1998 (first entry)
XX
DE Bacillus stearothermophilus S-layer protein sbas-A.
XX
KW S-layer; sbas-A; vaccine; adjuvant; carrier; hybridisation assay;
KW molecular spinning nozzle; molecular laser.
XX
OS Bacillus stearothermophilus.
XX
FH Key Location/Qualifiers
FT Peptide 1..30
FT /label= sig_peptide
FT Peptide 31..1228
FT /label= mat_peptide
XX
PN DEL9603649-A1.
XX
XX
PD 07-AUG-1997.
XX
PF 01-FEB-1996; 96DE-1003649.

PR 01-FEB-1996; 96DE-1003649.
 XX (LUBI/) LUBITZ W.
 PA (SLEY/) SLEYTR U.
 XX
 PI Kuen B, Lubitz W, Sleytr U;
 XX
 DR WPI; 1997-394558/37.
 DR N-PSDB; AAT75487.
 XX
 PT Preparation of S-layer proteins by expressing sbs-A gene in Gram
 PT negative bacterium or new sbs-B gene in any host, also new
 PT recombinant proteins containing heterologous inserts, e.g.
 PT epitope(s), useful as vaccines and adjuvants
 XX
 PS
 XX Claim 1; Pages 9-14; 31pp; German.
 CC The present sequence is the Bacillus stearothermophilus PV72
 CC S-layer protein, sbs-A. S-layer structures can be used as vaccines
 CC or adjuvants, particularly when they include a bacterial ghost that
 CC may contain additional epitopes in its membrane. Other uses of
 CC recombinant sbs-A, depending on the nature of the inserted peptide,
 CC are as an universal carrier for biotinylated reactants for use in
 CC immunological or hybridisation assays (the insert is streptavidin),
 CC to induce immune responses (epitopes), as a reagent for removing
 CC cytokine or toxin from serum (antigenic epitopes), as a molecular
 CC spinning nozzle (polyhydroxybutyrate synthase) and as a molecular
 CC laser (luciferase).
 XX
 SO Sequence 1228 AA.

Query Match 6.4%; Score 248.5; DB 18; Length 1228;
 Best Local Similarity 22.4%; Pred. No. 1.3e-05;
 Matches 201; Conservative 109; Mismatches 309; Indels 279; Gaps 44;
 QY 55 NLRIDDAKSFADSOQWTFPIAAVEKAGVIRKGTNGCEPRKIDRVMSALIVATK 114
 DB 323 dvalstgdkltvdas---tpientleykvvvkylk---dkngkefkeadafklinda 375
 QY 115 LDRKVGNGTPAT-----KFKLETLN-----WGK-----137
 DB 376 vvgvqvgntvntmtsvnaagftddtlvrvfdklapetvnsnvtldvetgkripv 435
 QY 138 -----EKANILVELGISVGTGDQWE---PKRYTV--KAEAQPIAKTDKQFGTEAAKVES 187
 DB 436 iastgsftlftlkealvtgkykylainnvkltlgynaeayelv-----ftanaaptv 489
 QY 188 AKAVT-----QKVEKFSKAVKRLKREDIKVT 215
 DB 490 ataptlgtltsltsltnvvgklagvnaagltypglqfttfackldestladnflv 549
 QY 216 NKANNDKVLKEVTLSEDKSATVELYSNLAQOTYVDNKKVGTPE--VAVGSLEAKTI 273
 DB 550 vkesgtvvaselkynadadakmvtlvpkadlkentlgyiklkglsksgleigtvneky 609
 QY 274 EMADQTVVADEPTALOFTVKD-ENGTEFVSPGIEFVTYPAEKINAKGEITLAKGST-- 330
 DB 610 eftqgdltl--pvlsvtskngdaglkvt--eageftvkfesenltnfnatvsgstltyg 665
 QY 331 -----TYKANVKKDG-----KVVAESKE-----348
 DB 666 qvavkvaganlsalatsadlilasveavtgdgtlykvkvaanqlerngykllvfgkgata 725
 QY 349 -VWVSAEGAIVASISMTVAEONKADFTSKDFKONKNVGVGDNNAVVEVKLDQENAVTGG 407
 DB 726 pvdadaanaatlacnylftfttegg-dvtlaptv---tkvifgsg-----lkda-davtl- 773
 QY 408 KVEYESINTEVAVVAKATGKVTYLSAGKAPVYTV-----KDSKGRALVSHTEIEFAFA 461
 DB 774 -----ltnvadag---kftfigfseelktssg--slvgskvtvekl 809
 QY 462 QKAMKDIKLEKTVALSTKQVTDLTKVAPVLDQYGKEFTAPVYTKVLDDKGKELKECKLE 521

DB 810 nngwvd---agltvsv-----apktlangkvtaavvltglidndkxak-lliv 856
 QY 522 AKVYNRELVNMAAGQ-----EAGNYTVVLTAKSGEKEKAKATTLALTKAPGAFSK 570
 DB 857 vdksscdgldadvagnvlkekdillirynswrhvasvkaadkdq-----nasafp- 908
 QY 571 FEVRGLDTELD---KYVTEBNOKNAAVTSVLPVDANGVLKGAEAALKYTTTNNKEGEV 627
 DB 909 -----tstaidtkslivne-----tdlaevkpenlvkvdagaavagfvtaldgs-- 956
 QY 628 DARDQVTVONNSVITVGGAKAGETFKYTVVDG-----KLITTHSKKVYDFAP 677
 DB 957 -----tnkvtfipsqelkagltvsvrl--dgvrdkvgnltiskylt--stkvtsanp 1003
 QY 678 TAKLAVEFTSTLSKEVAPADILKALNLSVDGVP-ATTAKATASNYEFVSAD-TNVV 735
 DB 1004 t-----lsslsiadgavnvdrsktltlefdsypnphtlckkdgsftnylvnn 1055
 QY 736 AENGT---VGAKGAT--SIYVKNLTVVKDK-----EQKVEFDKAVOVAVSIKAKPA 783
 DB 1056 nenktykivfhkgvltldftqyelaavskdftqtdidskvrlf---ltgsvaldevkpa 1110

RESULT 8
 ID AAW93252 standard; Protein; 1228 AA.
 XX
 AC AAW93252;
 XX
 DT 24-AUG-1999 (first entry)
 XX
 DE B stearothermophilus sbsa protein.
 XX
 KW sbsa protein; S-layer protein; Gram-negative; prokaryotic host cell;
 KW integration; cytoplasmic membrane; secretion; periplasmic space; toxin;
 KW eukaryotic host cell; vaccine; adjuvant; immunogenic epitope; luciferase;
 KW immunostimulant; cytokine; polyhydroxybutyrate; PHB synthase; body fluid;
 KW molecular laser; universal carrier molecule; monomolecular layer.
 XX
 OS Bacillus stearothermophilus.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..30
 FT Protein /label= signal_peptide
 FT /note= "sbsa"
 XX
 PN DEL9732829-A1.
 XX
 PD 04-FEB-1999.
 XX
 PF 30-JUL-1997; 97DE-1032829.
 XX
 PR 30-JUL-1997; 97DE-1032829.
 XX
 PA (LUBI/) LUBITZ W.
 XX
 PI Lubitz W, Resch S;
 XX
 DR WPI; 1999-122189/11.
 DR N-PSDB; AAX22748.
 XX
 PT Producing S-layer proteins in Gram-negative bacteria or eukaryotes -
 PT integrated into membranes or organelles or secreted into periplasma
 PT or growth medium, and nucleic acid encoding S-layer proteins with
 PT peptide insertions, used in vaccines or for enzymatic reactions
 XX
 PS Disclosure; Page 14-18; 34pp; German.
 CC This invention describes a method for the production of a S-layer protein
 CC (I) which comprises (a) preparing a Gram-negative prokaryotic host cell
 CC transformed with nucleic acid (II) encoding (I), linked to a signal

Query Match	6.48;	Score 248.5;	DB 20;	Length 1228;
Best Local Similarly	22.48;	Pred. No. 1.3e-05;		
Matches 201; Conservative	109;	Mismatches 309;		
		Totals	379;	Coverage 44

OY	55	NLPIDKOA	KPSEFADSOGOWTPPEFLAVERAGVIGKTGNGFPENPGINDRVSMSLLVEAYK	114
Dd	323	dvsistdgkklitvdas	-----tpefennetaykvvvbkjlk--dkngyefkedafteffrlinda	375
OY	115	LDTFYVNCGPAT	-----KFKNLELTN-----WGR----	137
Dd	376	vltvgvfgtnvcnncsvlnaaagftfdvddclttvwfcklllapevtvnssnvltldveylgkrlipv		435
OY	138	----EKANILVELGISVGTGDQWE---	PKRVTV---KAQAQFIARKTDPQGFTEAAKES	187
Dd	436	laatsgstfltlilkealvtgkykjalnnvktlctgyneaeyelv-----flanaspavl		489
OY	188	AKAATTT-----	OKVEVKRSKAVEKTLTKEDIKYT	215
Dd	490	ataptcltlygtllstcgsjltnwqglaggvneagtlypplgftttlatkldesltladntflv		549
OY	216	NKANANDVLVKEVTLSEDKRSATVELYSNLAKOTYTVDVNRKVGTKE--VAVSGLEARTI		273
Dd	550	veksgsvlvaseelkynadakmvltyphadlkentilygkikikjgksdgieltgvnektly		609
OY	274	EMADQTVVADEPTALQFTVVD-ENGTELVSPSEEGIEFTVPAAEKINAKEITLANKSTG-		330
Dd	610	eiftqbdlta--plvisvtskngdagikvlt--eaegfeivkkfsenIntfnattvsgtilyg		665
OY	331	-----TWKVAYYKKRGD-----KWVASKE-----		348
Dd	666	qvavpvkeganlsalstaadllpasvyeavetbgdgtykvkvaanqlerngyklivfsgkgata		725
OY	349	-VAVSABGAVAVASISMTFTVABQNKADFTSKDFECONKRVYEGDNARYVOVELKDQFNNAVTTG		407
Dd	726	pvtadaanaantlenylytfltlegv-dvtapdv---tkvfkgds-----lkda-davlt-		773
OY	408	KVEYESINTREAVAVDKATGKVTYLASGAPVYKVV-----KDSKGALVSHYTEIEFA		461
Dd	774	-----ltvndaqg---kftligfseelkitrsg-slyggkvktvelkt		809
OY	462	OKAMKDIKLEKTVALSTKCDVTDLKVVPAPVLDYGKEFFTAIPVYKVLVDKOGKELKEOKE		521
Dd	810	nngwvd--agbtctvsv-----aprtdeangkrtaavvlttgtidmnkdak-lrlv		856
OY	522	AKVNVNELVNLAGO-----EAGNYTVLTRAASKSEKAKATLALDELKAPGAFSK		570
Dd	857	vdhsacsdgcladvagvnikeddlilrynswhvltvasvkaadkdgg-----nasaaftp-		908

```

0Y 571 FEVRBLDLELD---KYTTEENQKRAMTVSVLPDPDANGVLKGAELAELKVTNTTNEGKEV 627
Db 909 -----tsaititktsllivefne-----tdlaevkpenivvkdgaagavagvtcalds-- 956
0Y 628 DATDAQVVOVONNSIYTGOGAKAGETPKYKVVLDG-----KLITTHSEKVVNDPAP 677
Db 957 -----chktvtfpsqelkagvtvsvlt--dgvrdrkvgnltisyytl--sfkvtvsanp 1003
0Y 678 TAKGLAVEFTSTSLKEVAPNADLKAAALLNLLSVDPG--ATTAKATASNVEFVSAD--TNVY 735
Db 1004 t-----lssistadgavnvdrsktltliefdsvnpmtlcllkadgsfnyllnyvn 1055
0Y 736 AENGT---VCAGKAT--SIYVKMLTVYKDGK-----EQVEFEDKAOYAVSIREAKPA 783
Db 1056 nenltkylivfhkgytltidefcygelavnskdtqtdtidsvtf-----itsvadevka 110

```

RESULT	9
AAB10625	
ID	AAB10625 standard; Protein; 1228 AA

AC AAB10625;

DT 08-JAN-2001 (first entry)

DE B. stearotheophilus sbsA protein.

KW S-layer gene; sbsA; antibacterial; vaccine; adjuvant; bioreactor;
KW polyhydroxyalkanoate synthesis.

05 *Bacillus stearotheophilus*.

	Location/Qualifiers
FH Key	1..30
FT Peptide	/label= signal_peptide
FT	31..1228
FT Peptide	/label= mature_peptide

PN DE19903345-A1.

PD 03-AUG-2000.

PF 28-JAN-1999; 99DE-1003345.

PR 28-JAN-1999; 99DE-1003345.

PA (LUBI/) LUBITZ W.

PI Lubitz W;

DR WPI; 2000-533868/49.

DR N-PSDB; AAA71797.

DR N-PSDB; AAA71797.

PT	Host cell, useful e.g. as bioreactor for production of poly(hydroxyalkanoate), containing two or more recombinant polypeptides, with at least one in carrier-bound form -
PT	
PT	

PS Disclosure; Page 12-16; 26pp; German.

This invention describes a novel host cell (A) comprising at least two functional recombinant polypeptides (I), at least one being in carrier bound form. The products of the invention have antibacterial activity. (A), or, where bacterial, their ghosts (B), are useful as vaccines or adjuvants (specifically for presentation of immunogenic epitopes of pathogens or autologous immunostimulatory polypeptides, e.g. cytokines), or preferably, as enzyme reactors for performing a cascade of reactions, specifically synthesis of poly(hydroxyalkanoate). Localization of individual (I), specifically enzymes, in separate cellular compartments avoids adverse reactions between products and substrates, when being used as bioreactors. (I) can be produced in carrier-bound form without loss of function. This sequence represents the *Bacillus stearothermophilus* S-layer protein sbSa which is used to illustrate the method of the invention.

[illegible]

```

DT 15-JAN-1997 (first entry)
XX Haemophilus adhesion protein HA2.
DE Haemophilus adhesion protein HA2.
XX
XX Haemophilus adhesion protein; HA2: hsf protein; vaccine.
KM
XX
XX Haemophilus influenzae type b strain C54.
OS
XX
XX WO630519-A1.
PN
XX
PD 03-OCT-1996.
XX
XX 22-MAR-1996; 96WO-US04031.
XX
XX 24-MAR-1995; 95US-0409995.
XX
XX (UWSL-) UNIV ST LOUIS.
XX (UNIW ) UNIV WASHINGTON.
XX
XX Barenkamp SJ, St Geme JW;
PI
XX MPI; 1996-455364/45.
DR N-PSDB; AAT41476.
XX
XX Recombinant Haemophilus adhesion proteins HA1 and HA2 - for use in
PT vaccines against H. influenzae infection.
XX
XX
PS Claim 5; Page 66-73; 120pp; English.
XX
XX Haemophilus adhesion protein HA2 (AAR9393) is associated with the
CC formation of surface fibrils involved in adhesion to various host
CC cells; it is also referred to hsf (haemophilus surface fibrils).
CC Its amino acid sequence was deduced from a genomic DNA clone
CC (AAT41476) derived from Haemophilus influenzae type b strain C65.
CC Large quantities of recombinant HA2 can be produced in transformed
CC prokaryotic or eukaryotic host cells, for use in vaccines against
CC H. influenzae infection.
XX
XX
SQ Sequence 2353 AA;

```

QY 389 DNAYVOVELKDQENAVTTGKVEYESLNTFVAAYVDKATGKTV-----LSGKAPVKYTK 443
DB 455 etv-----tfkagknlvkqgdganftyslgdaltgtsltgltngndakvlnk 506
QY 444 D-----SKGKALVSHTVLEIAFAOKA-MKDILKLEKTNYVALSKVDYDLKVKAPVLDQY 495
DB 507 dglitltpagnngtctgtntltstvtkdglkagnka1-----tnvasglrayd----- 551
QY 496 GKFTAPVTVKVLDDKQKELKEQ-----KLEAKYVNR-LVLMAGQEGAGNTYV-- 543
DB 552 -----anfvlmsatdlrnhvedaykgllhneknankgplvtstatavgdldklkg 604
QY 544 -VLTAKSGEKEA--KATLALELKAPGAFSKFEVRGLDELDKYVTEENOKNAMTVSVLPV 600
DB 605 vvvstkngtkeesngvqgadevltfga-----gaatvtsk---senghltltsvaet 654
QY 601 DANGVLKGAEMAEIKVTTNKKEGKVDATDAQVTVONNSVITVGGCAKGEFTYKVVVL 660
DB 655 kacgglekdgdltlkxvndgn-----cdnvltyvgn----- 685
QY 661 DGLITTHSFKVVDTAPTAAGLAVEFTSTSLKEVAPN-ADLKAALNLTSDGVPATTAK 719
DB 686 -gtavtktggtetvktgat-----dadrgkvtkdatandadkkvatvkdv-----atain 734
QY 720 ATASNVEFVSADTNVVAENGSTVGAK-----GATSIYV--KNLTVYKDGKEQKVEDKAV 771
DB 735 saatfvtktenlttsidednptdngkdaalkagdlitfkagknlvkvrdrkg--nltfdlak 792
QY 772 QVAVSIKEAK 781
DB 793 n--levktak 800

RESULT 12
AAB23860
ID AAB23860 standard; Protein: 2411 AA.
XX AAB23860;
AC AAB23860;
XX
DT 17-JAN-2001 (first entry)
XX
DE Haemophilus influenzae adhesin (Hia) protein from type c strain API.
XX
KW Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;
KW non-typable Haemophilus influenzae; antiinflammatory; audiotory;
KW antibacterial; meningitis; epiglottitis; septicemia; otitis media;
KW diagnosis; immunogenic; antigen.
XX
OS Haemophilus influenzae.
XX
PN WO200055191-A2.
XX
PD 21-SEP-2000.
XX
PF 16-MAR-2000; 2000MO-CA00289.
XX
PR 16-MAR-1999; 990S-0268347.
XX
PR (CONN-) CONNAUGHT LAB LTD.
XX
PA Loosmore SM, Yang Y, Klein MH;
XX
PI WPI; 2000-61897/59.
XX
DR N-PSDB; AAA92459.
XX
PT Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for
PT use as antigens and vaccines and for treating Hemophilus influenzae
PT infection -
XX
PS Claim 1; Fig 24; 275pp; English.
XX
CC The present sequence represents a Haemophilus influenzae adhesin
CC (Hia) protein from the type c Haemophilus influenzae strain API.

CC Hia genes and proteins have antiinflammatory, auditory and antibacterial
CC activities, and can be used in the production of a vaccine. An
CC immunogenic composition comprising an Hia gene, a polypeptide encoded
CC by an Hia gene, or a recombinant Hia polypeptide is useful for inducing
CC protection against disease caused by Haemophilus strains in a
CC susceptible host, preferably a human. An Hia protein is useful as an
CC antigen, in immunogenic preparations including vaccines, as a carrier
CC for other immunogens, and in the generation of diagnostic reagents. Hia
CC is useful for treating diseases caused by the infection of Haemophilus
CC influenzae such as meningitis, epiglottitis, septicemia and otitis
CC media. Recombinant production of Hia favours high recovery of the
CC protein compared to the low recovery of native protein from Haemophilus
CC influenzae species. A truncated protein has a significantly higher
CC amount of recovery than a full-length protein.
XX
SQ Sequence 2411 AA;

Query Match 6.1%; Score 238.5; DB 21; Length 2411;
Best Local Similarity 23.0%; Pred. No. 0.00011;
Matches 195; Conservative 102; Mismatches 333; Indels 219; Gaps 42;

QY 23 EKGAIVGNDKGMFEPKELTRAEATMMQILNLPIDKDK-PSFADSGQWTPPIA-A 80
DB 79 ekevtenmgwlyfhnkgyllkagaitlkagd-nlkxqstnaassfyslkxldltlsva 137
QY 81 VERAGV-----IKGTGNGFE---PNGKI-----BRVMSASILVAVRYKIDFK 118
DB 138 teklsfngdkvdltsdangllaktngnynhngldstlpavcnhtglsstfnd 197
QY 119 VNGTPATKFKDLEFLWNGKKEKANIIVELGISVGTGDMQEPKTVYKAEOQTAKTDKOF 178
DB 198 vektraetvkdvlmagvnlkgakt-----agvvesvdvlvsaynveflltgdntll 248
QY 179 GTEAAKVESAKAVTTQKVVEKFAVEKLEKEDIKV--TNKANND--KVLKVEYLSDDR 235
DB 249 dvlvtakengkt-----tevkftprktsvlkekdgklltfgemndntvntatcdndeg 303
QY 236 SATVELYSNLAQKOTYTVDNKVG---KTEVA-----VGSLEAKTTEMDQTVAD 283
DB 304 nglvtakavida-----vnkagrvvlttlangngdftvasgntvtesgdgtas- 355
QY 284 EPTALOPTVVDENGTEVSP-----EGIEFVTPAAEKINAKGETTLAKTSTTVKAVYK 337
DB 356 -----vltkdngnltvkvdydkvgdglkf--dsdckl-----vadftaltvlg-- 396
QY 338 KDKGVAAESKE---VKVSAEGAIVASISNWTVAEONKADPTSKD-----FKONNKVEGD 389
DB 397 --gkvaeiakeedkklvmagdlvtalgnlswkakaead--tdcdgalegyskdekvake 453
QY 390 NAVVOVELKDQFNNAVTTGKVEYESLNTFVAAYVDKATGKTV-----LSAKKAPVKYTKD 444
DB 454 tv-----tfkagknlvkqgdganftyslgdaltgtsltgltngndakvlnkd 505
QY 445 -----SKGKALVSHTVLEIAFAOKA-MKDILKLEKTNYVALSKVDYDLKVKAPVLDQY 496
DB 506 gltltpagnngtctgtntltstvtkdglkagnka1-----tnvasglrayd----- 549
QY 497 KEFTAPVTVKVLDDKQKELKEQ-----KLEAKYVNR-LVLMAGQEGAGNTYV--- 543
DB 550 -----anfvlmsatdlrnhvedaykgllhneknankgplvtstatavgdldklkg 603
QY 544 VLTAKSGEKEA--KATLALELKAPGAFSKFEVRGLDELDKYVTEENOKNAMTVSVLPV 601
DB 604 vvvstkngtkeesngvqgadevltfga-----gaatvtsk---senghltltsvaet 653
QY 602 ANGLVYLKGAEMAEIKVTTNKKEGKVDATDAQVTVONNSVITVGGCAKGEFTYKVVLD 661
DB 654 adsgglekdgdltlkxvndgn-----cdnvltyvgn----- 683
QY 662 GKLTTHSFKVVDTAPTAAGLAVEFTSTSLKEVAPN-ADLKAALNLTSDGVPATTAKA 720
DB 684 gtavtktggtetvktgat-----dadrgkvtkdatandadkkvatvkdv-----atains 733

```

OY 721 TASNVEFVSADTNVAVNGFTVGAK-----GATSIYV--KNLTVVKDGRFQKVEFPDQAVQ 772
D 734 aatfvkcentitsidednpdngkdalkegdltftfkaglnlvkrdgk--nlftdlakn 791
OY 773 VAVSIRKEAK 781
D 792 --Levktak 798

RESULT 13
AAR12083
ID AAR12083 standard; Protein; 1116 AA.
XX
AC AAR12083;
XX
DT 01-AUG-1991 (first entry)
XX
DE HWP protein.
XX
KM HWP; protein; food.
XX
OS Bacillus brevis HPD 31.
XX
FH Key Location/Qualifiers
FT Peptide 1..53
FT /label= sig_peptide 31..53
FT /label= sig_peptide 54..1116
FT Protein /label= mat_protein

JP03094683-A.
PN
PD 19-APR-1991.
XX
PF 06-SEP-1989; 89JP-0229304.
XX
PR 06-SEP-1989; 89JP-0229304.
XX
PA (HIGE-) HIGETA SHOU KK.
DR WPI: 1991-159801/22.
DR N-PSDB; AAO11789.
XX
PT New protein HWP gene - comprising specified DNA sequences,
PS permitting produ. of large ams. of HWP for use as food protein
XX
PS Disclosure; Fig 1; 8pp; Japanese.
XX
CC Using the sequence encoding this protein, HWP can be produced in
CC large amounts by recombinant techniques. The protein is used in food.
CC The first amino acid is encoded by the triplet TTC, according to
CC the specification.
XX
SO Sequence 1116 AA:

Query Match 6.0%; Score 231.5; DB 12; Length 1116;
Best Local Similarity 21.6%; Pred. No. 0.0001;
Matches 193; Conservative 125; Mismatches 307; Indels 269; Gaps 47;
OY 17 SINYLVKGAIVKGNKDKMFEPRKELTPRAEATMMAOILNPIIDKDAK-----PSFADSOQ 71
D 70 tvkrleelglvagyngdffgadtlttraetaltivargl--egaklaqfntlycdvrs 127
OY 72 -QWTPPIAAVEKAGYIKG--TGNGFEPNCKIDRVSMASLVEAYKLDTVVNGTPATKFKD 129
D 128 tdwfaagfvnvaageelkygfpdksfkpqngvtyaaavnlvralygepsvegy----- 180
OY 130 LETLNMCKEKANILVELGISVGTGDOMEPRKITYTKAAEAOFIAKTIDKQFGTEAKVESAK 189
D 181 -----vpsnmslqsgseiniakg-----lmpnmqgfaatlfr----- 212

```

```

OY 190 AVTTOKEVNEFSKAVEKLTREDIKVTNK---ANNDKVLVKEVTLSEDK--RSATVELSN 244
D 213 -mldnalrvkllmeqleqtdirlnvtdetlltkylkvevrimdwahegnshelpvtn 271
OY 245 LAAQOTYVDVKNKVGTEAVNGSLEAKTTEMAADQTV--VAD--BPTA-----LQFTVKDE 295
D 272 vpaiglgslkanew-----clngkadlgsnltkykvaeglnpnafdggkygvwlkdd 323
OY 296 NGTEVVSPEGIEFVTPAAEKINA--KGEITLAKGTSTTVKAVYKKGDKVAAESKEYKVS 352
D 324 renvlyvmegsededvmdrvalslylkgkaf-----tddivvdlksd-----ldvkt 373
OY 353 AEGAANVASISNWTVAEONK-----ADFTSKDFKONN---KVEGDN--AYVO 394
D 374 mdg-----seksyrltedtklcyntfrfndpvdalsklykdnldfgykvvlnndneveylh 429
OY 395 VELKDQFNAVTTGKVEYESLNTTEVAVVDKATGKTVTVLSACK----- 435
D 430 l-iddqtdkdvkgykys--kvlskidakkkltnlndskfsdledqdegskdflvflg 486
OY 436 APKVY-TVKDS-----KGAIVSHTVEIEAFQAQAMKDILEKTNVALSTKDVY 483
D 487 gpaikgdlkesdvsvyadgdkkyliv-----fanrvaeagkyek---vvsrntc 534
OY 484 DLKVKAPVLDQYGRKEFTAPVTVKVLDDKGEKLEKOKIEAKTVNR--ELVINAAGQENAGY 541
D 535 dir-----lvgyk-----lykvy-pdaasyenankdvkkvnsdlilslndgeev--- 579
OY 542 TVVITLAKSGEKAATLALELKAPGAF-----SKFEVRGIDPE 579
D 580 kllldpsgrvrhietkdaiddrklplaltkgtatynsktdydtfvtmqgktivslidqk 639
OY 580 --LDKYVTEENOKRAMVSVLPVDANGLVLKGAELAEIKYTTTNKSGKEVDATDAQTVQ 637
D 640 diyrivnyvdknsdk-----rgafekdlvellqpkvkvksedsatdangtv- 684
OY 638 NNSVITVGGAGKAGETTYKVVVLDGKL-----IT--THSFYV 672
D 685 ---llevnfsk-gevdkv-vldsklkysektwkladeddavvgdyevdaktavfkm 739
OY 673 V-DTAP---TAKGLAVEFTSLSEKVAAPNADT-----KAALNILSVDGVPATPAK 719
D 740 tgdilpatlgtklrgelknahtakfkdvakksdklkwysvdedkgevgalflvvgd-----s 793
OY 720 ATASNVEFVSADTNVAVNGFTVGAKGATSIYVKNLTVYKDG-----KEQKVEFD 768
D 794 glgqdhqf-----gmvkqyglaskgqdtlti-----vtkdgsdvekeykldgd 836

RESULT 14
AAB15945
ID AAB15945 standard; Protein; 2383 AA.
XX
AC AAB15945;
XX
DT 05-OCT-2000 (first entry)
XX
DE E. coli proliferation associated protein sequence SEQ ID NO:302.
XX
KM Escherichia coli; E. coli; proliferation; inhibition; screening;
XX antimicrobial; bacterial growth; antisense therapy; antibacterial.
XX
OS Escherichia coli.
XX
PN W0200044906-A2.
XX
PD 03-AUG-2000.
XX
PF 27-JAN-2000; 2000WO-US02200.
XX
PR 27-JAN-1999; 99US-0117405.

```


CC strains Joyce, K1, K21, LCC02, PMH1, 15 and 12. The nucleic acids and
CC vectors are used for the production of recombinant H. Influenzae HMW
CC proteins which can be used as vaccines to mediate a humoral or
CC cell-mediated immune response to provide protection against diseases in
CC humans caused by H. Influenzae (e.g., otitis media, epiglottitis,
CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC antigens in immunoassays for detecting antibodies against Haemophilus,
CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC HMW proteins can be used to isolate and clone hmw genes from other
CC non-typeable strains of Haemophilus via hybridisation reactions. The
CC present sequence represents a mature HMWA protein from a non-typeable
CC strain of H. Influenzae.
xx
SO Sequence 1095 AA;

Query Match 5.8%; Score 224; DB 21; Length 1095;
Best Local Similarity 19.6%; Pred. No. 0.00025;
Matches 157; Conservative 129; Mismatches 306; Indels 210; Gaps 37;

```
OY 69 SOGQWTPPIAAVERKAGVTKGTNGFEPN-----GKIDRVSM-----SLVEAY 113
DB 345 segstratf--tlesdltlnatgnlslnqvagidgnlqkslvanknltfegnltlaad 402
OY 114 KLDTKVNCGTPATK---FKDLETNNKGEKANILVELGIS---VGTGDQWEPKKTVTKA 165
DB 403 kkleleikgnltvkeganvltisaanygndksalslrgnvlknltvtsalnleknltve 462
OY 166 EAAOFIATKTDQOFTEA-----AKVESAKAVTTQKVEVFKSKAVEKLTKEDI----- 212
DB 463 gsakflaanpnyfinsvgldfnqgksnlsakgyahfk-dlnltxslnltltsdsayrtli 521
OY 213 --KVTNKANNDKVLKVEYTLSEDKRSATVELYSNLAQK--TYTVDNKRVG-----KTE 262
DB 522 egnltn-angd-----lnltdknmaelqignlsqkegnlttsdsknltntnqltkkg 574
OY 263 VAVGSLEAKTTEMADQTVVADEPTALOFT---VKDENGTEVSPGIEFVTAAEKINA 318
DB 575 vnkedsdsstannanlilkke--lqltgdlnlsqfkaeltakegadlilignsdnnn 631
OY 319 KGEITLAKGSTTTVKAAVYKKQGVVAKESKEVSAEGAAVASISMWTVAEQNKADFTS-- 376
DB 632 -----anakkvtfngv--kdsksaadhvnlhnsk---vetsngndaesngdqtalt 680
OY 377 --KDFKONNKVYEGDNAYVOVELKDQFNA-----VTTGKVEYESLNTVEV-AVVDK 423
DB 681 lnaknltvnnhl---tshkcvnltsasenvtkagtltnatgsyevtakgtgdlkgykes 736
OY 424 ATGKVTYLSAGKAPV-----KTVVKDSKGKALVSHTEIEAF---AQKAMKDILKLEK 472
DB 737 tsgsvltlatgelaavsnlsgntvlitankglltqagsvtasalngvtasssgdlsqtl 796
OY 473 TNVALSTKDVTDLKAKAVLDOYGEFTAPTVKVLDDGKELKQKLEAKYVNRLEVLN 532
DB 797 sgnltvksaialdltcks-----gselkaktgean-----vls 828
OY 533 AAGQFACNYT---VVLTAKSGEKEAKATLALELKAPAFSKEFEVGLDTELDKYVTEENQ 589
DB 829 atgltgltlsgnavntantgd----- 850
OY 590 KNAMTV-SVLPVDANGLVLKGAEEAELEKVTNTNKEGKVDATDAQVTVONNSV-ITVGOG 647
DB 851 --ltveedaakidatg---ga-----atltatagkltlkassaltsaannqvnlsakdg 897
OY 648 AKAGETVYVTVVLDGKLLTTHSFKVVVDNAPPAKGLAVEFTSTSLKEVAPNADLKALLNI 707
DB 898 slggnlnaanvln---ltgal-----tlvkgsslnansgclvinakdaelngeasgn 947
OY 708 LSVGVFPATTAATASANEFEVSADTNVAE---NG--TVGAKGATSIYVKNLTV---V 757
DB 948 hlv--vnatnangsgsvlatcsrvnlgtldltnnglnlisknglnvlylkgvkldvkyi 1005
OY 758 KDGKEQKVEFDKAVOVAVSIKE 779
```

DB 1006 qpgiasvdevleakrilekvkd 1027

Search completed: March 23, 2002, 07:51:23
Job time: 3521 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2002, 06:54:37 ; Search time 61.69 Seconds

(without alignments)
286,352 Million cell updates/sec

Title: US-09-754-947-1

Perfect score: 3885
Sequence: 1 AGKTPDPVADHWGIDSINY.....EPDKAVQVAVSIKKAAPATK 785

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	378.5	9.7	1222	2	US-08-682-517-15 Sequence 15, Appl
2	378.5	9.7	1252	4	US-08-682-517-9 Sequence 9, Appl
3	240.5	6.2	2353	4	US-09-377-155-33 Sequence 33, Appl
4	240.5	6.2	2353	4	US-08-913-942-4 Sequence 4, Appl
5	238	6.1	1912	3	US-08-409-995-4 Sequence 4, Appl
6	238	6.1	1912	3	US-08-685-467-4 Sequence 4, Appl
7	216.5	5.6	1529	2	US-08-728-470-10 Sequence 10, Appl
8	216.5	5.6	1529	4	US-08-719-641-10 Sequence 10, Appl
9	214.5	5.5	1600	2	US-08-617-697-10 Sequence 10, Appl
10	214	5.5	1338	2	US-08-728-470-9 Sequence 9, Appl
11	214	5.5	1338	4	US-08-719-641-9 Sequence 9, Appl
12	214	5.5	1477	1	US-08-682-517-9 Sequence 9, Appl
13	214	5.5	1477	1	US-08-302-832-4 Sequence 4, Appl
14	214	5.5	1477	2	US-08-530-198-4 Sequence 4, Appl
15	214	5.5	1477	2	US-08-469-880-4 Sequence 4, Appl
16	214	5.5	1477	2	US-08-728-470-4 Sequence 4, Appl
17	214	5.5	1477	2	US-08-617-697-4 Sequence 4, Appl
18	214	5.5	1477	2	US-08-719-641-4 Sequence 4, Appl
19	214	5.5	1599	2	US-08-617-697-9 Sequence 9, Appl
20	213.5	5.5	1183	2	US-08-447-031K-2 Sequence 2, Appl
21	212.5	5.5	1098	1	US-08-409-995-2 Sequence 2, Appl
22	212.5	5.5	1098	3	US-08-685-467-2 Sequence 2, Appl
23	212.5	5.5	1098	4	US-09-377-155-32 Sequence 32, Appl
24	212.5	5.5	1098	4	US-08-913-942-2 Sequence 2, Appl
25	207	5.3	1561	3	US-08-894-017-23 Sequence 23, Appl
26	207	5.3	1536	1	US-08-682-517-15 Sequence 2, Appl
27	207	5.3	1536	1	US-08-302-832-2 Sequence 2, Appl

28	207	5.3	1536	2	US-08-530-198-2	Sequence 2, Appl
29	207	5.3	1536	2	US-08-469-880-2	Sequence 2, Appl
30	207	5.3	1536	2	US-08-728-470-2	Sequence 2, Appl
31	207	5.3	1536	2	US-08-617-697-2	Sequence 2, Appl
32	207	5.3	1536	2	US-08-719-641-2	Sequence 2, Appl
33	200.5	5.2	1565	6	5352450-2	Patent No. 5352450
34	188	4.8	2123	4	US-08-968-685A-10	Sequence 10, Appl
35	186	4.8	679	4	US-08-913-942-15	Sequence 15, Appl
36	183.5	4.7	198	2	US-08-682-517-19	Sequence 19, Appl
37	180	4.6	1026	2	US-08-614-377A-7	Sequence 7, Appl
38	180	4.6	1026	4	US-09-142-648B-7	Sequence 7, Appl
39	176	4.5	1026	4	US-08-194-290-7	Sequence 7, Appl
40	175.5	4.5	573	4	US-08-235-836C-112	Sequence 112, App
41	174	4.5	200	2	US-08-682-517-20	Sequence 20, Appl
42	174	4.5	658	1	US-08-409-995-5	Sequence 5, Appl
43	174	4.5	658	3	US-08-685-467-5	Sequence 5, Appl
44	174	4.5	658	4	US-08-913-942-5	Sequence 5, Appl
45	172.5	4.4	1545	4	US-08-296-791-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-682-517-15
; Sequence 15, Application US/08682517
; Patent No. 5874267
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/682,517
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ. ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1222 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-682-517-15

Query Match 9.7%; Score 378.5; DB 2; Length 1222;
Best local Similarity 22.1%; Pred. No. 4.2e-18;
Matches 223; Conservative 129; Mismatches 378; Indels 279; Gaps 44;

QY	16	DSINLYEKGAVGNDKMGKPEPKELTRAEATIMMAOILPLDIDAKPFPAD-SOGOWY 74
DB	15	EAVALDQDVOIGDNGNPNPLNTVYRAQAEIFTKALELNGDV--EFKVKAGAWY 72
QY	75	TPFLAAVERKGVKIG-TGNCFEPPNGKIDRVSMASLVEAKLDTKVGNGTATPKFEDLETL 133
DB	73	YNSIAAVANGIEBGSATFAPNKSILTRSEAAKILVEAFGLGEGAD--LSEFADASQV 129
QY	134	N-WGKEKANILVELGISVGT-GDQWEPKKTVTYKAAQFIATKT-DKQGTAEAKVESAKA 190
DB	130	KPAKAKYLEIAVANGIEFGTDANKLNPNNSITRQDFALVFRRTVDKEGFTPEAAFWKA 189
QY	191	VTTQKVEVRSKRAVEKLT-----EDIKVTN---KANDKVLV-----KEVTLSED 233
DB	190	INNTVEVTEETEEVTNVALNFKIEGLETKMASVKQNRKVVALLTTEAQTADDEYVLTLD 249
QY	234	KRSATVELYSNLAAKQTYTDVANKVGKTEVAV--GSLEAKTIEADQTVVADDEPTA--LQ 289
DB	250	--GETTIGCFGVAA-----VPTKVELYSVAVGKLGQEVKQAKTYVAAGQSKAGIPVT 302

```

QY 290 FIVADNGTEVYS-----PEGIEFTTPAA-----313
D 303 FIVPGNNNDGVPTLTGALNTNEGATATSYTRYKEGTDDEVTAATGDRSKFSLGIVFWG 362
QY 314 -EKINAGEIT-----LAKGSTTYKAVYK--KDGK-----VAESKEVY 351
D 363 VDTILSVEETTTASVNNNGANKTKYTKTKPKTGPEAKNKTFFNNGVEENNVISDKYANA 422
QY 352 SARGAASVATISNKTVAEQNKADFTSKDKRKNNKYEGDNAYQVELKD-----Q 400
D 423 TVNGVAKALQJLSTNGTALD--AAQITDSKGGEATFFVSGTNAAVTPVYVDLHSTNNSTNKK 480
QY 401 FNA-----VTTGKREYESLNTAEVAY-----VDATGKVITYLSGKA--PKVYTKVKSCKALY 451
D 481 YKSAALQTTASKYVTFALDAEITIELTRADNA--GEVAAGATNGREKVIYVKDKAGNLAK 539
QY 452 SHRVEIE-----AFAOKAMDKILEKTN-----474
D 540 NEIVNAFNEEDKDRVISTVTNAKFEVDPTDPYAVFTQDPAKAKQISV--KTNQKGEATPYIGS 598
QY 475 -----VALSTRKQVTDLKYKAPVLDQYGKEFTAPVT--VKVLDKDGKELKEQKLEAKRYN 526
D 599 DTVNDYATPTIAMIDINTSDAQGDIDEEGEPKAVAPISIFQAPYLDGSAIKAYK---KSDL 655
QY 527 RELVLNAAQEGAGNTYVVLTAKSGEKEAKATLAELEKAPGAFSFEVRGDETFELVYNE 586
D 656 NKAVTKFQDSEETAVFAEALVNGSGKKVYGIS-----KKATYITYNGANDIKVDNQVIS 710
QY 587 ENQKNAMTVSVLPVANGLVYKGAENAE-----LKVTYTT---NKEGKVDATDA 632
D 711 PNNSTVITYEA--TLSTGTGVTTPAKNLEVTYVDGKTTAVVIVATGIAVNTDGDYAFYAK 769
QY 633 QVTV-----ONNSVITVGO-----GAKAGEPYK-----655
D 770 EATATFTATNEVPNSYTGATQPTNADSGSNSINSIWAGKNPKYKYGVSCKTKYFGANG 829
QY 656 -----VTVVLDKLLTTHSFKRVVDPTAPTAGLAVFEFTSL 691
D 830 NEVFGEAMAEALLTOYATGEGOKVTISYVNDGDTVT--FKVIS-----AVNSTEAI 878
QY 692 KEVANP---ADLKAALLILSVSDVPAPTTAKATASNVEFVSADPTNYA---ENGVIYAGK 745
D 879 KVAVTPPTAAPTGTGLTITTPAAGSLVDLTATATNTLGLISLADADLNASATVTPATVATSLKD 938
QY 746 ATSIYVKKLVYVKDGKEQKEFEEDKAVO-----VAVSIEKEAKPA 783
D 939 SANNSL-SLTLVETGANTGV--FATTVQAGTLISLTAGTLIVTAYDAKNA 985

RESULT 2
US-08-682-517-9
: Sequence 9, Application US/08682517
: Patent No. 5874267
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: Expression of surface layer proteins
: NUMBER OF SEQUENCES: 25
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/682,517
: FILING DATE:
: CLASSIFICATION:
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1252 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-682-517-9

```

```

Query Match      9.78; Score 378.5; DB 2; Length 1252;
Best Local Similarity 22.18; Pred. No. 4.4e-18;
Matches 223; Conservative 129; Mismatches 378; Indels 279; Gaps 44;

QY 16 DSIYLVKCAVKNDKGFEPGKELTRAAATMMQIINLPIDKCAKSPAD--SGGMY 74
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 45 EAVQALVDQVIGQDTNGNPNPLNTYTRQAALIFPKALELANGV--NFDVRAGAY 102
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 75 TPFLAAVEKAGVIGK--TGNGFEPNGKIDRYSMASLLVEAKYLDTKYNGTPARKFDLET 133
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 103 YNSIAAVVANGIEGVSATFEAPENKSLTSEAKILVEAFGLEGEAD--LSEPDASGY 159
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 134 N-MCKEKANLLVLEGLSVGF--GDOMEPKTYTKAAEAQIAT--DKQFTEAAKVESAKA 190
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 160 KPAKKKLETLAIVANGIEGTDAKKLNPNSITRQDPALEFKRTYDKVEGETPEEAPFYKA 219
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 191 VTQKVEKPSKAVKELTK-----EDIKYTN--K-KNNDKVLY-----KEYTLSPD 233
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 220 INNTYVEEPPEEVTVQALNFKIEBELKIKMSVYKQTKKKVYVLTTEAQATADKEVYLTLD 279
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 234 KRATVEELVSNLAAKQTYVDVAKGVKTEVAV--GSLKATIMADQTVVADEPTA--LQ 289
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 280 --GETIGGFEGGVA-----VPTKELVSAVQKIGQGVKQVQAIVTVAEGSKAGIPVT 332
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 290 FTVKDENGTEVVS-----PEGIEFTVPRA----- 313
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 333 FTVPGANNNGVVPFLTGLALTNDEGIATSYTRYKEGIDEVTAIYATGDSKSFSLCYFWG 392
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 314 -EKINAKGELT--LAKGTSTTKAVYK--KDGKV-----VAESKEVY 351
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 393 VDTILSEVEETGASVNNQANKTYKYTKYKPNPTGKPEAKNTEPVGVENMNTVSDKVANA 452
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 352 SAEGAIVASISMTVAEQKADPTSDFRONKKVYESGNAYQVVELKD-----Q 400
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 453 TVNGVVALDLSNGTALD--AAQITDSKGEATFYSGTNAATPVVYDLHSTNNSTSKK 510
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 401 FNA-----VTTGKVEESLNTTEVAV--VDKATGKVTVLSAGKA--PYKVTAKDSKALV 451
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 511 YSASALQTTASKVTEPAALQAEYTIELTRADNA--GEVAIIGATNGREKVIYIKDAAGNLAK 569
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 452 SHVYLE-----APAQAMMDIKLEKTN----- 474
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 570 NELIVNAEFEDKDRVISTVTNAKFEVDTPDAVYTTGDKAIOIS--KTNDKGEATFYIGS 628
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 475 -----VALSTKDVTDLKVKAPVLDQYGEFTAPVT--VKVLDDKDGELKROKLEAYVN 526
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 629 DTVNDYAPPIAMIDINTSDAKOGDLDEBPRAVAPISIFQAPYLDGSIKIRAYK--KSDL 685
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 527 RELVLTMAAQEQENGNTVVLIAKSGEKAATLALTELARGAFSFEYRGILDELDKRYTE 586
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 686 NNAVTRKFDGSEITAVFAELVNMOSGKRVGTISI-----KKATYTYLNTGANDIKVDNQVIS 740
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 587 ENOKNAMTVSVLPVDPANGLVLKGAFAE-----LAKYTT-----KKEGKVATDA 632
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 741 PKRSTVIYEA--TLSSITGVITPAKNLEWISVDCKTAAVKVIAIGLVNMDGDQYAFPAK 799
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 633 QVTV-----ONNSVITYQ-----GAKAGEYTK----- 655
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 800 EATATFTATNEVPNSYTGATQFMADSGSINSIMWAGKNPVYAGVCSKTYTKYGGANG 859
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 656 -----VTVVLDKGLITTHSEKVVVDTPAPAKGLAVEFTSTSL 691
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 860 NEVFGEAAMEALLTOYATEGOKVTISYVNDGDTV--FKVIS-----AVNSSTEAI 908
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 692 KEVAPN---ADLKAALLIISVDGVPTTKAKATASNEEFSADPTNYA---ENGTYGANG 745
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 909 KPVAPTTPAATFTGALLILTPAAGGLVLTATNTLGISLADADLNSATTVDPATVATSLDK 968
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 746 ATSIYVKNLVVKKDGKQKVEFDKAVO-----VAVSITKEAKPA 783
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 969 SANNSL--SLITVETGANTGV--FATTVAGTSSSITAGCTTIVYATDAKKNA 1015
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```
RESULT 3
US-09-377-155-33
; Sequence 33, Application US/09377155
; Patent No. 6197312
; GENERAL INFORMATION:
; APPLICANT: PEAK, Ian Richard Anselm
; APPLICANT: JENNINGS, Michael Paul
; APPLICANT: MOXON, E. Richard
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0128
; CURRENT APPLICATION NUMBER: US/09/377,155
; CURRENT FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 33
; LENGTH: 2353
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-377-155-33
```

```
Query Match 6.2%; Score 240.5; DB 4; Length 2353;
Best Local Similarity 22.7%; Pred. No. 4.5e-08;
Matches 193; Conservative 102; Mismatches 336; Indels 219; Gaps 41;
```

```

23 EKGAVKNDKMEPEKELTRAEATMAOILNPIOKDA-----KPSFADSGQWTPF 77
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
79 EKEVTENSNNGIYDNGKVLKAGAITLKAGD-NLKIKONTDESTNASSFTSLKKDLTL 137
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
78 IA-AVEKAGV-----IKGTGNGFE---PNCKI-----DRVSASLIVK 114
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
138 TSVAETKLSGANGDKYDITSDANGKLAKTNGNVLHNGIDSTLPDAVNTFGLSSSF 197
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
115 LDRVNGTPTAKFKDELTLNNGEKEKANILVELGISVGTQDMEPEKTYTKAEAAQFIKT 174
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
198 TPNDEKTRATATYKVDVNLACGNIKAGT-----AGGVESEVDLVSAANNFEITGD 248
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
175 DKQGTAAKAVESKAKAVTTQKVEYKSKAVEKLTREKDIKVTNRAND--KVLVEKVTLS 231
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
249 KNTLDVVLTKAKNGKT-----TEVKFTPKTSVIEKEDGKLTGKENDTNKVTSTNTATDN 303
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
232 EDKRSATVELYSNLAAGOTYVDVNVKYG--KTEVA-----VGSLEAKTIEMADQT 279
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
304 TDBNGGLVTKAKAVIDA-----VNKAGWRVKTJTANGONGDFATVAVSGTNVTEESGDGT 356
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
280 VVADEPTALQFTVDENGTEVYSP-----EGIEFVTPAEKINAKGEITLAKGSTPTVK 333
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
357 TAS-----VTKDTNGNGITVYKDAKVGDLKF--DSDDKI-----VAODTALTALT 399
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
334 AVYKKGDKVAESKE--VKVSAEGAAVASISN--WTVAEONKADFTSKDEKONNKYEG 388
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
400 G-----GKVAEIAKEDDKRKLVNAGDLVTALGNLSMRKAKADPDGALLEGISQDEYKAG 454
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
389 DNAYVQVELADQFNAVYTTGKVEYSLNTEVAVVDKATGKVTY-----LSAGKAVKVTYK 443
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
455 ETV-----TRKAGNKLKVKODGANFTYSLQDALTGLSITLIGGTTNGGNDKATYINK 506
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
444 D-----SKGKALVSHTVLEIAFAOKA-MKDIKLEKTNVALSTKDVYDLKVKAPVLDQY 495
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
507 DGLITTPAGNGGTGTWTISTVKDGIKAGNKAI-----TNVASGLRAVDD----- 551
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
496 GKERTAVYTKVLDKDKKELEKQ-----KLEAKTYVNR-LVLNAGOEAGNTYV-- 543
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
552 -----ANPDVNLNSATDLNRHVEDAYKGLDNLNERNKANKQPLVTDSTATAVGDRLKLG 604
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
544 -VLTAKSGEKEA--KATLAELEKAPGAFSKFEVGLDPELDKYVTEENOKAMVSVLPV 600
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```

DB 605 WYVSTKNGTKESNOVKADEVLFTGA-----GAATVYSK-----SENGKFTITVSAET 654
OY 601 DANGVLKGAEAELKVTYTNKEGKEVDATDAQVTVONNSYITVGQAKAGETKYVTVL 660
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 655 KADGKLEKDGDTIKLKQDNQ-----TDNVLTVGN----- 685
OY 661 DGKLTTHSPRVVDTATPAKGLAVEFTSTSLKEVAPN-ADLKALNLNLSIDGVPATTAK 719
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 686 -GTAVTKGFEETVKTGAT---DADRGKVTYKDATANDADKRVATVADV-----ATAIN 734
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 720 ATASNVEFSADTVNVAENGTVGAK-----GATSIYV--KNLTVVKDGKQKVEFPKAV 771
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 735 SAATFVKTENLTISIDEDNPTDNGKDALKAGDILTRKAGNKLKVKRDKG--NITFDLAK 792
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 772 QVAVSIREAK 781
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 793 N-LEVKTAK 800
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```

RESULT 4
US-08-913-942-4
; Sequence 4, Application US/08913942
; Patent No. 6200578
; GENERAL INFORMATION:
; APPLICANT: St. Geme, Joseph
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/913,942
; FILING DATE: 29-DEC-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/4031
; FILING DATE: 22-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Vance, Dolly A.
; REGISTRATION NUMBER: 39,054
; REFERENCE/DOCKET NUMBER: A-61053-1/RFT/RMS/DAV
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2353 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-913-942-4
```

```

Query Match 6.2%; Score 240.5; DB 4; Length 2353;
Best Local Similarity 22.7%; Pred. No. 4.5e-08;
Matches 193; Conservative 102; Mismatches 336; Indels 219; Gaps 41;

23 EKGAVKNDKMEPEKELTRAEATMAOILNPIOKDA-----KPSFADSGQWTPF 77
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```

Db 79 EKEVTENSMGIVFDNKGVLKAGAITLKAGD-NLKIKONTDESTINASSFTYSLKKDLTDL 137
OY 78 IA-AVERKAGV-----IKGTGNFE-----PNKGI-----DRVSNASLIVEAYK 114
Db 138 TSVATEKTSFPAANGDKVITSDANGKLAKTGNVHNLGDLSTLPDAVTGTVLSSSSF 197
OY 115 LDTKVNKGTPTATKFKDLETLNNGKEKANIIVELGISVGTGDDMEPKKTYTKAEAOFIKT 174
Db 198 TPNDVEKTRAATVYKDVLAAGNNIKGAKT-----AGNVESVDLVSAVNNVEFTIGD 248
OY 175 DKQFTEAKVESAKAVTTQKVEVFSKAVEKTEKEDIKVTNKNAND-KVLVEKTVLS 231
Db 249 KNTLDVVLTAENKGT-----TEVFTPKTSVIEKEKDKLFTGKKNNDTKVTSNTATDN 303
OY 232 EDKRSATVELYSNLAOKOTYVDNKGK-----KTEVA-----VGSLEAKTITEMADOT 279
Db 304 TDEGNGLVTAFAVIDA-----VNKAGMRVKTTFANGONGDFATVAGTNTVFESGGGT 356
OY 280 VVADPTALQFTVKDENGTEVVS-----EGIEFVTPAERKINAKGEITLAKGSTTVK 333
Db 357 TAS-----VTKDTNGNGITVKIDAKVGDGLKFDSDKFI-----VADTTALITVT 399
OY 334 AVYKKDGKVVAAESKE---VKVSAEGAAVASISN---WTVAEQNKADFTSKDFQNNKVIK 388
Db 400 G-----GKVAELIAKEDDKKILVNAAGDLVTALGNLSMKAKAEADTDGALGISKDOEVKXG 454
OY 389 DNAYVOVELKDQFNNAVTTGKVEYESLNTVEAVYVDKATGKVTV-----LSAGKAPVKTVK 443
Db 455 ETV-----TFKAGKNLKVADGKANFTYSLQDALITGLTITLGGTTNGGDAKTIVINK 506
OY 444 D-----SKGALVSHVIEIAFAOKA-MKDIKLEKTVALSTKDVTDLKVAPVLDQY 495
Db 507 DDLTTPAGNGGTGTGNTISVTIKDKIKAGNKAI---TNVAGLIRAYDD-----551
OY 496 GKEFTAPVTVKVLDKDCKELKEQ-----KLEAKYVNR-LVINAAGQEAAGNTTV- 543
Db 552 -----ANFDVLNNSATDLNRHVEDAYKGLNLNEKNANKOPLVTDSTATVYGDRLKLG 604
OY 544 -VUTAKSGEKEA-KATLTALELKAPGAFSKFEVYRGIDTELDKVTYBENOKNAMTVSLPV 600
Db 605 WYVSTKNGTKREESNOVKOADEVLFTGA-----GAATYTSK---SENGHITTVSAET 654
OY 601 DANGLVKAEGAELKVTYTNKEGEVDAITDAQVTVQNNNSVITVGGACAAGETTKYTVL 660
Db 655 KADCGLEKGDGTIKLKYVDNQN-----TDNVLTVGNM-----685
OY 661 DCKLITTHSPKVVDTAPTAKGLAVEFTSLKEVAPN-ADLKAALINIISVDGVPATTAK 719
Db 686 -GTAATKGGEEYVTKCAT-----DADGKVTYKDATANDADKKVATYKDV-----ATAIN 734
OY 720 ATASNVEFVSADTNVVAENGTVGAK-----GATSIYV-KNLTYYKDGKEQKVEFDKAV 771
Db 735 SAATFVKTEULTISIDEDNTDNGKDALKAGDTLTFKAGKNLKVXRDKG--NITEDLAK 792
OY 772 QVAVSIKEAK 781
Db 793 N-LEVKTAK 800

```

```

OY 5
US-08-409-995-4
; Sequence 4, Application US/08409995
; Patent No. 5646259
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen I.
; APPLICANT: St. Geme III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adhesion Proteins
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehrt, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA

```

```

COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/409,995
FILING DATE: 24-MAR-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-61053/RFT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1912 amino acids
TYPE: amino acid
STRANDEDNESS: double
TOPOLOGY: unknown
US-08-409-995-4

```

```

Query Match      6.1%; Score 238; DB 1; Length 1912;
Best Local Similarity 22.8%; Pred. No. 56-08;
Matches 194; Conservative 101; Mismatches 339; Indels 216; Gaps 42;

```

```

OY 12 HMGIDSINTLVKGAVK-----GNDKGMFEPKGEITLRAEAAVT-MMAQILNLPIDKD 61
Db 86 NMGI-----YFDNKGVLKAGAITLKAGDNLKXQXTDEXTNASSFTYSLKKDLTDLVSAT 141
OY 62 AKPSFADSGOWTPTPIAAVERKGVK---GTGNG-PEPNG-----KIDRYSMASLIVEAYK 114
Db 142 EKISFPAANGDK-----VDITSDANGKLAKTGNVHNLGDLSTLPDAVTNTGVLSSSSF 196
OY 115 LDTKVNKGTPTATKFKDLETLNMGKEKANIIVELGISVGTGDDMEPKKTYTKAEAOFIKT 174
Db 197 TPNDVEKTRAATVYKDVLAAGNNIKGAKT-----AGNVESVDLVSAVNNVEFTIGD 247
OY 175 DKQFTEAKVESAKAVTTQKVEVFSKAVEKTEKEDIKVTNKNAND-KVLVEKTVLS 231
Db 248 KNTLDVVLTAENKGT-----TEVFTPKTSVIEKEKDKLFTGKKNNDTKVTSNTATDN 302
OY 232 EDKRSATVELYSNLAOKOTYVDNKGK-----KTEVA-----VGSLEAKTITEMADOT 279
Db 303 TDEGNGLVTAFAVIDA-----VNKAGMRVKTTFANGONGDFATVAGTNTVFESGGGT 355
OY 280 VVADPTALQFTVKDENGTEVVS-----EGIEFVTPAERKINAKGEITLAKGSTTVK 333
Db 357 TAS-----VTKDTNGNGITVKIDAKVGDGLKFDSDKFI-----VADTTALITVT 399
OY 334 AVYKKDGKVVAAESKE---VKVSAEGAAVASISN---WTVAEQNKADFTSKDFQNNKVIK 388
Db 400 G-----GKVAELIAKEDDKKILVNAAGDLVTALGNLSMKAKAEADTDGALGISKDOEVKXG 453
OY 389 DNAYVOVELKDQFNNAVTTGKVEYESLNTVEAVYVDKATGKVTV-----LSAGKAPVKTVK 443
Db 455 ETV-----TFKAGKNLKVADGKANFTYSLQDALITGLTITLGGTTNGGDAKTIVINK 505
OY 444 D-----SKGALVSHVIEIAFAOKA-MKDIKLEKTVALSTKDVTDLKVAPVLDQY 495
Db 507 DDLTTPAGNGGTGTGNTISVTIKDKIKAGNKAI---TNVAGLIRAYDD-----550
OY 496 GKEFTAPVTVKVLDKDCKELKEQ-----KLEAKYVNR-LVINAAGQEAAGNTTV- 543
Db 552 -----ANFDVLNNSATDLNRHVEDAYKGLNLNEKNANKOPLVTDSTATVYGDRLKLG 603
OY 544 -VUTAKSGEKEA-KATLTALELKAPGAFSKFEVYRGIDTELDKVTYBENOKNAMTVSLPV 600

```



```

Db 604 WVSTKNGTKEESNOVKQADAEVLFTGA-----GAATVTSK-----SENGKHTITVSAET 653
Qy 601 DANGLVJLKAFAELKVTYTNKEGKEVDATDAQYTVONNSVITVGQAKAGETVKVTVL 660
Db 654 KADCGLEKDDGTILKVDNQN-----TDNVLTGNN-----684
Qy 661 DGKLTTHSFVVDYAPLAKGLAVEFTSTLSKEVAPN-ADLKAALLNLSVGVPAATTAK 719
Db 685 -GTAVTKGFEFTVKTGAT-----DADRGKVTVKDATANDADKKVATVKDV-----ATAIN 733
Qy 720 ATASNVEFVSADTVNVAENGTVGAK-----GATSIYV--KNLVYKDGKQKVEFEKAV 771
Db 734 SAAFTVKTENLTSTIDEDNPTDNGKDALKAGDPLTFKAGKNLKVRRDGK--NITFDLAK 791
Qy 772 QVAVSIRKAK 781
Db 792 N--LEVKTAK 799

RESULT 6
US-08-685-467-4
; Sequence 4, Application US/08685467
; Patent No. 6060059
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohnbach, Test, Albritton & Herbert
; STREET: Four Embardero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,467
; FILING DATE: 22-JUL-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1912 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-685-467-4

Query Match 6.18; Score 238; DB 3; Length 1912;
Best Local Similarity 22.88; Pred. No. 5e-08;
Matches 194; Conservative 101; Mismatches 339; Indels 216; Gaps 42;
Qy 12 HMGIDSNVLEKGVK-----GNDKGMFEPKELTREAAT--MMAQILNLPDKD 61
Db 86 NMGI-----YFDNKGVLKAGAITLAKAGDNLKKKQATDEXTNASSFTYLSLKKRLDTLTSVAT 141

```

```

Qy 62 AKPSFADSGQWYTPFIAAVEKAGVYK--GTGNG-FEPNG-----KIDRVSNASILLVEAYK 114
Db 142 EKLSPGANGDK-----VDITSDANGLKLAKTGNVHLNGLDSTLPDAVTVTGVLSSSSF 196
Qy 115 LDTKYNGLPAPKRFDELTLNNGKEKANIYELGIVSGTGDMEPKYTKLEAAOFIKT 174
Db 197 TPNDVEKTRAATVADVLNAGNNIKGAKT-----AGCVNESVDLSAANNFEITGD 247
Qy 175 DKQFTEAKAVESAKAVYTKQVEYKFSKAVELKREDIKY--TKRANN--TVLYKEVTL 231
Db 248 KNTLDVULTAKEKNT-----TEVKFTPKTSVIEKEDKLTGKENNDTNKVTSTATDN 302
Qy 232 EDKRSATVELYSNLAQOTYTVDNKVG--KTEVA-----VGSLEKTIEMADOT 279
Db 303 TDEGNGLVAKAVIDA-----VNKAGMRYKTTANGONGDFATVAGSTVTFESGDGT 355
Qy 280 VVADEPTALOFTVDENGTEVYSP-----EGIEFVTYPAEKINAKGKITLAKGISTYK 333
Db 356 TAS-----VTKDTNGNGITVYKDAKVGDGLAF--DSDKI-----VADTTALTVT 398
Qy 334 AVYKKGKVAESKE---VKVSAEGAASISN--WVAEQNKADFTSKDKQNNKYEG 388
Db 399 G-----GKVAELAKEDDKKLVNAGDLVTALGNLSMKAKAEADTDGALGSKQDEYKAG 453
Qy 389 DNAYVVELKDQFNAVTTGKVEYESLNTFVAVVDKATGKVTY----LSAKAPYKVTYK 443
Db 454 ETV-----TFKAGKNLKVQDGANFTYSIQDALTGLTSLTGTTNGSNDAKTVYK 505
Qy 444 D-----SKKALVSHTEVELEAFQAKA--MKDIKLEKTNVALSTVDYDLVYKAPVLDQY 495
Db 506 DGLITTPAGNGTGTNTISVTKDGIKAGNKAI-----TNVAGSLRAYD-----550
Qy 496 GKEFTAPVTVKVDKDGKELKEQ-----KLEAKYVNE-LVLAAGQEGAGNYTV--543
Db 551 -----ANPDVLNNSATDLNRHVEDAKGGLLNLEKANNQPLVTDSTAFVGLRLKG 603
Qy 544 -VLAKSGEKA--KATLLELKAPGAFSKFEVRGLDTLKYYTEENQNNMTVSVLPV 600
Db 604 WVSTKNGTKEESNOVKQADAEVLFTGA-----GAATVTSK-----SENGKHTITVSAET 653
Qy 601 DANGLVJLKAFAELKVTYTNKEGKEVDATDAQYTVONNSVITVGQAKAGETVKVTVL 660
Db 654 KADCGLEKDDGTILKVDNQN-----TDNVLTGNN-----684
Qy 661 DGKLTTHSFVVDYAPLAKGLAVEFTSTLSKEVAPN-ADLKAALLNLSVGVPAATTAK 719
Db 685 -GTAVTKGFEFTVKTGAT-----DADRGKVTVKDATANDADKKVATVKDV-----ATAIN 733
Qy 720 ATASNVEFVSADTVNVAENGTVGAK-----GATSIYV--KNLVYKDGKQKVEFEKAV 771
Db 734 SAAFTVKTENLTSTIDEDNPTDNGKDALKAGDPLTFKAGKNLKVRRDGK--NITFDLAK 791
Qy 772 QVAVSIRKAK 781
Db 792 N--LEVKTAK 799

RESULT 7
US-08-728-470-10
; Sequence 10, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Matlaire, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia

```

```

: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent in Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/728.470
: FILING DATE:
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302.832
: FILING DATE: 16-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US PCT/US93/02166
: FILING DATE: 16-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9205704.1
: FILING DATE: 16-MAR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Berkstresser, Jerry W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-633
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1529 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-728-470-10

```

```

Query Match      5.6%; Score 216.5; DB 2; Length 1529;
Best Local Similarity 22.5%; Pred. No. 1.1e-06;
Matches 174; Conservative 109; Mismatches 328; Indels 161; Gaps 37;

```

```

QY 89 CTGCGFENGKIDRVSMSLVEAKIDTKVNGPATKFKLETLNMQKEKANLIVEIGI 148
DB 769 GENSSSTITGININITNKANVTLOA---DTSNSNTGLKK---RTITLD---NISVEGNI 817
QY 149 SVGTGDDWEPRKKTVAKAAOAFIAKTKQFSTEAKYVESAKAVTQK---VEYKESKAVEK 206
DB 818 SL-TGANANINIGNSLIAE-----DSTFKGEASDNINITGTFNNGTANININIKGVYK 868
QY 207 LTKEDIK-----VTNKANDKVLK-----EVTLSDEKRSATVELYSNLAKO-T 250
DB 869 L-QGDINNKGGLNITTNASGTOKTIINGNITNEKGLNINIKADAEIIOIGINISQKEGN 927
QY 251 YTVQDNKVG-----KTEVAVGSLEAKTIMADQTVVADE-PTALOFTVQDENTEVVS 302
DB 928 LTISSDKYNITNOITIKGVEGGRSDSEAEANLITITKELKLAGDLNISFNKAELTA 987
QY 303 PEGIEFTVPAEKINAKG-EITLAKGSTTYKAVYKKDKYVAESKEVKSAGAASVSI 361
DB 988 KNGSOLITIGNASGNADAKKVTFRKVKDSKIST---DGHVYTLNSEVYKS-NGSSNAGN 1042
QY 362 SNMYVAEONKADFTSKDKRONKKYVEGDNAVVOVELKD--QFNATVTKGVEESLN---- 415
DB 1043 DNSTGCLTISAKDVYVNNVNTSHKTNISAAAGNVTTKGTITNA-TTGSVEVTAONGTILK 1101
QY 416 -----TEVAVVDKATGKTVTVLSAGKAPVNVTVKDSKALVSHTVEI 457
DB 1102 GNITSQNTVATENTLVTEENAVINATSGTYNI-STKTGDIKGIGIESGNANNITASGNT 1160
QY 458 EAFACKAMKDIKLEKTNAVLESTKDVTDLKVAPVLDQYKGEFTAPVTVKVLDDKGEKELKE 517
DB 1161 LKVSNTIGQDVTVPADACALTTTASSTISATG-----NANITTKTGIDNG----- 1206
QY 518 OKLEAKYVNRRLVLAAGOEAGN---TVVLTAKSKEKAKATVTLLELKAPAFSKFEYR 574

```

```

DB 1207 -KVSSSGSVTLVATGATLAVGNISGNTVTTTADSG-----KLTSTVGSTINGTNS----- 1256
QY 575 GLDFELDKYVTEENOK-----NMTVSVLPVANGVLKGAFAELK---VTTNKEG 624
DB 1257 -----VTTSSQSGDIEGTISGNTVNV-TASGDLITIGNSAKVEKKNAATLITAEISG 1306
QY 625 KEVDATDAQTVQVON-NSVITVGGAGAKAGETRYKVTVLD--GKLITTHSEKVVADAPPAKG 681
DB 1307 KLTQYTGSSITSSNGQTLTFLAKDSSIAGNINANAVNTLMTGTGLTTGDSK----- 1356
QY 682 LAVETSTSLKEVPANADLKAA-----LNLISVDGVPATTAKTATASVNEFVSND----- 731
DB 1357 --INATSGTLITINAKDALDAAAGSDRTVNAVATNASGSGNVTAK-TSSSVN-ITGDLMTI 1412
QY 732 --TNVVAENG--TYGAKGATSIYKNLNVKDGKEQKVEFPKAVOAVASIKE 779
DB 1413 NGLNITISEGNRYVRLKG-KEIDVK---YIOPGVAASVEVIEAKRVLEKVD 1460

```

```

RESULT 8
US-08-719-641-10
: Sequence 10, Application US/08719641
: Patent No. 6218141
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Mattare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent in Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/719.641
: FILING DATE:
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302.832
: FILING DATE: 16-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US PCT/US93/02166
: FILING DATE: 16-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9205704.1
: FILING DATE: 16-MAR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Berkstresser, Jerry W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-625
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1529 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-719-641-10

```

```

Query Match      5.6%; Score 216.5; DB 4; Length 1529;
Best Local Similarity 22.5%; Pred. No. 1.1e-06;

```


Oy 628 DATDAQTVON-NSVITVGAGAGETKYVVLVD--GKLITTHSFKYVDTPAKGLAV 684
 Db 1380 TQTSSTSSGOTTLLAKDSSIGNINAAVNTLTGTTLTTGDSK-----I 1427
 Oy 685 EFTSTSLKEVAPNDLKA-----LNLISVDGPATTAATASVNEFVSAD-----T 732
 Db 1428 NATSGTLTINAKDALDGAASGDRTVVATNATNASSGCVNTAK-TSSSVN-TIGDLNTINGL 1485
 Oy 733 NVVAENG--TVGAGATSIYVKNLTIVKDKGEQKVEFDKAVQVAVSIKE 779
 Db 1486 NIISNGRNVRLRG-KEIDVK---YIOPVAVSEVIEAKRVLEKVD 1530

RESULT 10
 US-08-728-470-9
 ; Sequence 9, Application US/08728470
 ; Patent No. 5928651
 ; GENERAL INFORMATION:
 ; APPLICANT: Barenkamp, Stephen J
 ; TITLE OF INVENTION: High Molecular Weight Surface Proteins
 ; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Shoemaker and Mattare, Ltd.
 ; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
 ; CITY: Arlington
 ; STATE: Virginia
 ; COUNTRY: U.S.A.
 ; ZIP: 22202-0286
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/728,470
 ; FILING DATE:
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/302,832
 ; FILING DATE: 16-MAR-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US PCT/US93/02166
 ; FILING DATE: 16-MAR-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: GB 9205704.1
 ; FILING DATE: 16-MAR-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Berkstresser, Jerry W
 ; REGISTRATION NUMBER: 22,651
 ; REFERENCE/DOCKET NUMBER: 1038-633
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 415-0810
 ; TELEFAX: (703) 415-0810
 ; INFORMATION FOR SEQ ID NO: 9:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1338 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-728-470-9

Query Match 5.58; Score 214; DB 2; Length 1338;

Best Local Similarity 20.48; Pred. No. 1.4e-06;
Matches 194; Conservative 139; Mismatches 361; Indels 258; Gaps 46;

Oy 5 PPVDPADHWGIDISINYLVEKGA---VKGNDKGMEPEPKELTRAE-----AATMAAO- 52
 Db 399 YRDGGRYVWVNTTLN--VTSGSKFNLSIDSTGSGSTGSIINAEINLGTTFPKAIFNINAG 456

Oy 53 -ILNLPIDKADPSPADSGOGWTPFLAAVEKAGVIGKTGNGFBNKRIDRVSNASLIVE 111
 Db 457 STANPSIKASIMPRKSNAN---VALFNEDISVSG---CGSVNFKLMSNSSNIQTPGVILK 510
 Oy 112 AYK-----LDTRKNGTPATKFKDLETLMWGGEKANILVELGISVGTGOMPEPKTYT 163
 Db 511 SONFNVSOGSTLNLKAEOSTETAFESIENDLNLNATGCGITIRO--VEGTDSR-----VN 562
 Oy 164 KAAAOFLAKTDQOGEFBAKVESAKAVTQKVEVKSFAVEKLFKEDIKYTNKANNDKV 223
 Db 563 KGVA-----AKKNITFFKGNITFFGSOKAT---EIKGVNTINKNTNATIRANPAENKSP 614
 Oy 224 -----LVKEVYLSDEKRSATVELYSNLAAKOTYVDV-----NKVGTEVAV 265
 Db 615 LNIAGNVINNGNLTTAGSIINIAGNLVFSKGANLQATINNTFVNAAGSPDNNGASNISIR 674
 Oy 266 GSLAKTI-EMADOTVVADEPTALQFYVK-----DEN-----GTEVVSPE 304
 Db 675 GGAKEFDINNTSSLNITNSDTTYRTTIKGNISKSGDLNIDKSDAEIQIGNISQKE 734
 Oy 305 GIEFVTPAEKINKGETTLAKGT---STVKAIVYKKDGKVAESKEVKSAGAAVAVS 360
 Db 735 G-NLITSSDKVNITNOITTIKAGVGGRSDSSEA---ENANLTIOTKELKAGD-----LN 785
 Oy 361 ISNMTVAEONKADFTSK-----DEKONKV-YEGDNAYVOVEL 397
 Db 786 ISGF-----NKAETAKNGSDLTIGNASGNADAKKYTFDKVKDSKISTDGHNTVNLSEV 840
 Oy 398 KDGFNAVITTK-----VEYSILNTEVAVVDKAGSKVVLGSKAPV 438
 Db 841 KTSNGSSNAGDNDSTGLTISAKDVTVNNNTVSHKTINISAAAGVNTTKREGITTINATGVS 900
 Oy 439 KVTYKDSKKA-LVSHFVEIEA---FAOKAMDKLEKTNTVALSTKDYT-----D 484
 Db 901 EYTAQNGTIKGNITSQNTVATENLVTTENAIVNATSGVINISPKTGDIKGISTSGN 960
 Oy 485 LKVA-----PYLDQYKEFT-----APVYKVLKDGKEIKE 517
 Db 961 VNITASGNTLKVSNITGQDVVTADAGALTITAGSTISATGNANITTKTDING----- 1015
 Oy 518 OKLEAKYVNRRLVLAQAQEGANY---TVVLTAKSGEKAATLLEKAPGAFSKFEVR 574
 Db 1016 -KVSSSGSVTLVATGATLAVGNISGNTVTTIAGSG---KLITVSGTNGTNS----- 1065
 Oy 575 GLDELDKRYVTEENOK-----NAMTVSVLPVDANGVLKGAFAELK---VTTNKEG 624
 Db 1066 -----VTTSSQSGDIEGTISGNTVNV--TASGDLTINSKAVEKKNAATLTLASG 1115
 Oy 625 KEVDATDAQTVON-NSVITVGAGAGETKYVVLVD--GKLITTHSFKYVDTPAKG 681
 Db 1116 KLTTQTGSSITSSNGOFTLLAKDSSIGNINAAVNTLTGTTLTTGDSK----- 1165
 Oy 682 LAVFTSTSLKEVAPNDLKA-----LNLISVDGPATTAATASVNEFVSAD----- 731
 Db 1166 --INATSGTLTINAKDALDGAASGDRTVVATNATNASSGCVNTAK-TSSSVN-TIGDLNTI 1221
 Oy 732 --TNVVAENG--TVGAGATSIYVKNLTIVKDKGEQKVEFDKAVQVAVSIKE 779
 Db 1222 NGLNISNGRNVRLRG-KEIDVK---YIOPVAVSEVIEAKRVLEKVD 1269

RESULT 11

US-08-719-641-9
; Sequence 9, Application US/08719641
; Patent No. 6218141

; GENERAL INFORMATION:
 ; APPLICANT: Barenkamp, Stephen J
 ; TITLE OF INVENTION: High Molecular Weight Surface Proteins
 ; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Shoemaker and Mattare, Ltd.
 ; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

```

: STREET: Bldg. 1
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/719,641
: FILING DATE:
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302,832
: FILING DATE: 16-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US PCT/US93/02166
: FILING DATE: 16-MAR-1993
: APPLICATION DATA:
: APPLICATION NUMBER: GB 9205704.1
: FILING DATE: 16-MAR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Berkstresser, Jerry W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-625
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1338 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: US-08-719-641-9

```

Query Match 5.5%; Score 214; DB 4; Length 1338;

Best local similarity 20.4%; Pred. No. 1.4e-06; Matches 194; Conservative 139; Mismatches 361; Indels 258; Gaps 46;

```

QY 5 FPDVPADHMGIDSIYLVKGA--VKGNDKGFEPGKELTRAE-----ATMMAQ- 52
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 399 YRDKGRYNAVTLN--VTSKSKFNLSIDSTGSGSTPSIRNMLGITFNKATFNLAQG 456
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 53 -ILNLPIDKAKSFADSGQWTPPLAAYEKAGVYKGTGNGFEPNGKIDRVSMASLIVE 111
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 457 STANFSIKASIMPEKSNAN--YALFNEEDISVSG--GGSVNFRLNASSNIOQPGVLIK 510
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 112 AYK-----LDTKVNGPFAFKFDELTLMWCKEKNIIIVELGISVGTDDQWPKKTYT 163
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 511 SQNFNVSGSGTLLWKAEGSTETAFSTENDLNATGSGNITIRO--VEGTDSR-----VN 562
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 164 KAAEAOFIATKDFQFTEAKVESAQAVTTQKVEKFSKAVEKLTKEIDIVTNKANNDKV 223
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 563 KGA---AKNITFKGNGITPFSQAKT---EIKGNVYINKNTNATLNGANFAEKSP 614
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 224 -----LVKEYTLEDKRSATVELYSNLAQKQTYVDV-----NKVQTEVAV 265
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 615 LNIAGVNNNGNLTTAGSIINAGNLTVSKGANLQAITNTYFNAGSFDNNGASNISIAI 674
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 266 GSI-EAKTI-EMAOQTVADPFLAQFTVK-----DEN-----GTEVVSPE 304
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 675 GGAKEFDINTTSLNITNTSDTYRTIIRKGNISNKSGLDILIDKSDAEIQIGNISQKE 734
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 305 GIEFVTPAEKINAKGEITLAKGT---STTVAVYKKDKKVAESKEVVSAGAAVAS 360
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 735 G--NLTISSDKVITNGITIKAGVEGGRSDSSER---ENNLIITQTELKLAGD---LN 785
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 361 ISMWVYAEONKADFTSK-----DFKONKQV-YEGDNAVYVOVEL 397
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

```

DB 786 ISGF-----NKAEITAKNSDGLTIGNASGNADAKVTPEKVDKSIISTDGHNVITLSEV 840
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 398 KDOFNAVTTGK-----VEYESINTEVAVVDKATGKVTYLSACKAPV 438
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 841 KTSNGSSNMGNDNSTGLTISAKDVTVNNNVTSHKTTINISAAAGNVTTKEGTTINATGVS 900
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 439 KVTYKDSKGA-LVSHITVEEA-----FAOKAMKDKLEKTINVALSTKQVT-----D 484
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 901 EYTAONGTIGKNTISQNVYATATENLVTEENAVINATSGVINISTGTGDKGIESTSGN 960
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 485 LKKA-----PVLDOYKKEPT-----APYVKVLDDQKELKE 517
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 961 VNITAGSNILKSNITIGQDVTYADAGALTTTGTISATFGNANITTKTGIDNG----- 1015
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 518 QKLEARYVNRVELYNAGQAGNY---TVVLTKSGSEKKAATLALAKFASKEFVR 574
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 1016 -KVESSSGSVTLATGATLAVNGISGNTVITADSG---KLSTVSGTINGTNS----- 1065
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 575 GLDTELDKVTYENOK-----NMTVSYPVDANGVLKGAELK---VTTNKEG 624
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 1066 -----VTTSSQSGDIEGTISGNTVNV-TASTGDLTIGNSAKYEAKNGAATTLAESE 1115
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 625 KEYDATDAQVTVON-NSVITVGGAKAGETVKTVYLD---GKLITHSFKYVDTAPYAKG 681
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 1116 KLTTQTGSSITSSNGQTTLAKDSSIGNINAAVGLNTTGLITIGDSK----- 1165
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 682 LAVEFTSLKEVAPNADLKAA-----LNLITLSDVGVPATTAKATASNYEVSAD---- 731
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 1166 --INATSGTLINAKAKLDGAASGRTVYNATNAGSSGVATK-TSSSVN-ITGDLNTI 1221
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 732 --TNVVAENG--TVGAKGATSIYKNIITVYKDKKEQKVEPKAVQVAVSIE 779
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 1222 NGNIISENGRNTVRLRG-KEIDVK---YIOPGVAHYEEVTEAKRYLEKVKD 1269
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

RESULT 12

US-08-038-682-4

; Sequence 4, Application US/08038682

; Patent No. 5549897

; GENERAL INFORMATION:

; APPLICANT: BARENKAMP, STEPHEN J

; APPLICANT: ST. GENE HII, JOSEPH W

; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS

; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Shoemaker and Mattare, Ltd

; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

; STREET: Bldg. 1

; CITY: Arlington

; STATE: Virginia

; COUNTRY: U.S.A.

; ZIP: 22202-0286

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/038,682

; FILING DATE: 16-MAR-1993

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: BERKSTRESSER, JERRY W

; REGISTRATION NUMBER: 22,651

; REFERENCE/DOCKET NUMBER: 1038-293

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 415-0810

; TELEFAX: (703) 415-0813

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1477 amino acids

; TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-038-682-4

Query Match 5.58; Score 214; DB 1; Length 1477;
Best Local Similarity 19.9%; Pred. No. 1.6e-06;
Matches 167; Conservative 122; Mismatches 311; Indels 238; Gaps 39;

2 GKTPPDVPAHMG-----IDSYIVLEKGA-----VGNDDKGMPEPKELTRAEAT 48
Db 748 GSVFPIYAHNSGGAELKMSKSEIN--ISNGANFTLNSHVAGDD--AFKINKDL-----T 797
QY 49 MMAQILNLPIDKDAKPSFADSGQWYPTFAVEKAGVINGKGTGNGFEPNGKIDRVSMAS- 107
Db 798 INATNSNFSL-RQTKDFDYDGYAR-----NAINSTYNISILGQNVTLGGQSSSSITGN 850
QY 108 -LVEAKKLDTKVNGTPATKFKDLETLMWKEKANILVELGISVGTDDQPEPKTYTAKAE 166
Db 851 ITTEKAAVNTLEANNAPNOQNIIRDVRILKG---SLVNGSLSL-TGENADIKGNLTISE 905
QY 167 AAOFIATDKQFGEAKKVESAKAVTTQKVEVFKSKAVEKL---TKEDIKVTNKANNKD 222
Db 906 SATFKGTRDTLNTITNGFTNNGTA-----EINTQGVYKLGNTVNDGDLITTHAKRNQ 959
QY 223 -----VLVEK--VLTSEDKRSATVELYSNLAQKQ--TYVDVKNKVGKT-EVAVGSLEAK 271
Db 960 RSIIGGDIINKKGSINLTSDNNDAEIQIGNISQKEGNLTISDKNITKQITIT---KK 1015
QY 272 TIEMADQTVVADEPTALOFTVKDENGTEVVSPEGIEFVTTPAAEKINAKGETTLAKGTSTT 331
Db 1016 GIDGEDSSDATSANANLTITKTELKLTEDLSISGFN-----KAEITAKDGRDLT 1064
QY 332 VKAVYKKDKGVVAESKEV-----KVSAGAAVASISMTVAEONKADFTSKDFQKN 383
Db 1065 IG--NSNDGNSGAELAKTYTFNNVYKDSKISADGHNVTLNSKYKTSSSNGGRESND----- 1117
QY 384 KYVEGDNAVVOVELKDOFNNAVTTGKVEESLNTFEVAVYDKATGKVTYLSAGKAPVYTVK 443
Db 1118 -----NDTGLTITAKNVEVKNKIDTSLKTVNITASBKVYTTAGS-----TIN 1158
QY 444 DSKGKALVSHVTEIEAFQAQAKMIKLEKTNVALSTKDVIDLAKKAVLDQYCEFTAPV 503
Db 1159 ATNFGKA-----SITTKTGDISG-----TISGNTVSVA 1186
QY 504 TVKVLDDGKELKBOKELEAKYVNRLEVLNAGOEAGNV---TVVLTAKSKEKEKATLAL 560
Db 1187 TVDLTTSGSGKIEAKSEAN-----VTSATGTIGGTISGNTVAVTANAG----- 1231
QY 561 ELKAPGAFSKFEVARGLDTELDKYVTEENQKAMTVSVLPVDANGVLKGAFAELKVT- 619
Db 1232 -----LTV-----GNGCAEINATEGATTLATG 1253
QY 620 ---TNKSGKEVDATDAQVY--QNNSVITYGQGKAGETKYVYLDGKLITTHSEFVND 674
Db 1254 NTLTTEAGSSITSTKGVYDLAOGST-----AGSINAAVNTLN-----TTG 1295
QY 675 TAAPTAKLAVEFTSTSL---KEVAPNADLK--AALLNITLSDVGPPTTAKATASNEFV 728
Db 1296 TLTTVAGSDIKATISGTLVYINAKDKLNGDASGDSDEVNANVASSSGSVTA-AISSSVN-1 1353
QY 729 SADTNVVAENGTVGAKGATSIYVKNLTVVKDKQEKVEFDKAVQVAA-SIKEAKPAK 785
Db 1354 TGDINTVNGNLIISKDRNIVRLR-----GKEIEVKY---IQPGVASYVEVTEAKR 1401

RESULT 13
US-08-302-832-4
Sequence 4, Application us/08302832
Patient No. 5603938
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J

TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/302,832
FILING DATE: 16-SEP-1994

CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992

PRIOR APPLICATION DATA: US pct/us93/02166
APPLICATION NUMBER: US pct/us93/02166

FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:

NAME: Berkstresser, Jerry M
REGISTRATION NUMBER: 22,651

REFERENCE/DOCKET NUMBER: 1038-404
TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0810

INFORMATION FOR SEQ. ID NO: 4:
SEQUENCE CHARACTERISTICS:

LENGTH: 1477 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
US-08-302-832-4

Query Match 5.58; Score 214; DB 1; Length 1477;
Best Local Similarity 19.9%; Pred. No. 1.6e-06;
Matches 167; Conservative 122; Mismatches 311; Indels 238; Gaps 39;

2 GKTPPDVPAHMG-----IDSYIVLEKGA-----VGNDDKGMPEPKELTRAEAT 48
Db 748 GSVFPIYAHNSGGAELKMSKSEIN--ISNGANFTLNSHVAGDD--AFKINKDL-----T 797
QY 49 MMAQILNLPIDKDAKPSFADSGQWYPTFAVEKAGVINGKGTGNGFEPNGKIDRVSMAS- 107
Db 798 INATNSNFSL-RQTKDFDYDGYAR-----NAINSTYNISILGQNVTLGGQSSSSITGN 850
QY 108 -LVEAKKLDTKVNGTPATKFKDLETLMWKEKANILVELGISVGTDDQPEPKTYTAKAE 166
Db 851 ITTEKAAVNTLEANNAPNOQNIIRDVRILKG---SLVNGSLSL-TGENADIKGNLTISE 905
QY 167 AAOFIATDKQFGEAKKVESAKAVTTQKVEVFKSKAVEKL---TKEDIKVTNKANNKD 222
Db 906 SATFKGTRDTLNTITNGFTNNGTA-----EINTQGVYKLGNTVNDGDLITTHAKRNQ 959
QY 223 -----VLVEK--VLTSEDKRSATVELYSNLAQKQ--TYVDVKNKVGKT-EVAVGSLEAK 271
Db 960 RSIIGGDIINKKGSINLTSDNNDAEIQIGNISQKEGNLTISDKNITKQITIT---KK 1015
QY 272 TIEMADQTVVADEPTALOFTVKDENGTEVVSPEGIEFVTTPAAEKINAKGETTLAKGTSTT 331
Db 1016 GIDGEDSSDATSANANLTITKTELKLTEDLSISGFN-----KAEITAKDGRDLT 1064
QY 332 VKAVYKKDKGVVAESKEV-----KVSAGAAVASISMTVAEONKADFTSKDFQKN 383
Db 1065 IG--NSNDGNSGAELAKTYTFNNVYKDSKISADGHNVTLNSKYKTSSSNGGRESND----- 1117

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2002, 06:57:02 ; Search time 66.49 Seconds

(without alignments)
899.339 Million cell updates/sec

Title: US-09-754-947-1

Perfect score: 3885

Sequence: 1 AGKTFPPVADHMGIDISINY.....EFDKAVOVAVSITKAKPATK 785

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database :

1: PIR:68:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3847	99.0	814	2 I40048	S-layer protein pr
2	760	19.6	874	2 JC4930	S-layer protein pr
3	445	11.5	1099	2 T14850	S-layer protein pr
4	402.5	10.4	1176	2 A33856	surface-layer 125k
5	279.5	7.2	404	2 F59097	hypothetical prote
6	268	6.9	1036	2 T30311	S-layer protein -
7	259	6.7	1109	2 A56143	surface-array prote
8	255.5	6.6	762	2 A34355	cell surface prote
9	248.5	6.4	1228	2 I40468	surface layer prote
10	245.5	6.3	1832	2 S33441	EF protein - Strep
11	241	6.2	1128	2 T30296	R27.2 protein - Tr
12	240	6.2	3468	2 T34418	hypothetical prote
13	237.5	6.1	652	2 B59102	hypothetical prote
14	236	6.1	745	2 B33856	hypothetical 80K p
15	234.5	6.0	1231	1 A48490	endo-1,4-beta-xyla
16	234.5	6.0	6642	2 T29757	protein UNC-89 - C
17	232.5	6.0	1087	1 S41797	cellulose 1,4-beta
18	232.5	6.0	1829	2 T32583	hypothetical prote
19	231.5	6.0	1116	2 A35129	surface layer prote
20	231	5.9	2383	2 D64962	extracellular matr
21	228.5	5.9	2055	2 T31110	endo-1,4-beta-xyla
22	220.5	5.7	1148	2 S72635	exo-poly-alpha-gal
23	218.5	5.6	1302	1 JC6009	surface-located me
24	217	5.6	1324	2 T18265	endo-1,3(4)-beta-g
25	216.5	5.6	1417	2 D85521	probable adhesin e
26	215.5	5.5	1234	2 S72640	endo-1,4-beta-xyla
27	214	5.5	1477	2 B43855	high-molecular-wel
28	213.5	5.5	1185	2 A42404	collagen adhesin -
29	212.5	5.5	1556	2 A60988	saliva-interactin

30	212	5.5	2139	2 T18296	myosin heavy chain
31	211.5	5.4	1365	2 T30832	limp1 protein - Myc
32	211	5.4	1561	1 S06839	surface antigen sp
33	210	5.4	2269	2 T28677	thoptry protein -
34	210	5.4	1315	2 T28679	fibrinogen-binding
35	207	5.3	1536	2 A43855	high-molecular-wel
36	205	5.3	837	2 JN0292	antigen 332 - mala
37	205	5.3	2660	2 E85822	probable invasin 2
38	204	5.3	1084	2 A28555	middle wall protei
39	203.5	5.2	2032	2 T39917	hypothetical prote
40	203	5.2	1361	2 T03415	S-layer protein -
41	201	5.2	1565	2 S04729	surface antigen pa
42	199.5	5.1	631	2 JC6031	scaffoldin dockeri
43	199.5	5.1	1983	2 G86643	hypothetical prote
44	199	5.1	941	2 S29043	cellulase (EC 3.2.
45	196	5.0	1503	2 T01098	chloroplast outer

ALIGNMENTS

RESULT 1	
I40048	S-layer protein precursor - Bacillus anthracis
C:Species: Bacillus anthracis	
C:Date: 02-Aug-1996 #sequence-revision 02-Aug-1996 #text-change 21-Jul-2000	
C:Accession: I40048; S51695	
R:Etienne-Toumeilin, I.; Siraard, J.C.; Duflot, E.; Mock, M.; Fouet, A.	
J. Bacteriol. 177, 614-620, 1995	
A:Title: Characterization of the Bacillus anthracis S-layer: cloning and sequencing o	
A:Reference number: I40048; MUID:95138020	
A:Accession: I40048	
A:Status: preliminary; translated from GB/EMBL/DBJ	
A:Molecule type: DNA	
A:Residues: 1-814 <RES>	
A:Cross-references: EMBL:Z36946; NID:q1359499; PIDN:CAA85408.1; PID:q634013	
C:Genetics:	
A:Gene: sap	
C:Superfamily: S-layer repeat homology	
F:34-88/Domain: S-layer repeat homology <SLR1>	
F:95-147/Domain: S-layer repeat homology <SLR2>	
F:156-209/Domain: S-layer repeat homology <SLR3>	
Query Match	99.0%; Score 3847; DB 2; Length 814;
Best Local Similarity	99.0%; Pred. No. 3.4e-154;
Matches 777; Conservative 4; Mismatches 4; Indels 0; Gaps 0;	
QY 1 AGKTFPPVADHMGIDISINYLVKGAAGVKGNGEPNGKIDRVASASLVAAYKLDFTVFN 120	
DB 30 AGKTFPPVADHMGIDISINYLVKGAAGVKGNGEPNGKIDRVASASLVAAYKLDFTVFN 120	
QY 61 DAKSPFDSOGQWTFPIAAVEKAGVKGNGEPNGKIDRVASASLVAAYKLDFTVFN 120	
DB 90 DAKSPFDSOGQWTFPIAAVEKAGVKGNGEPNGKIDRVASASLVAAYKLDFTVFN 149	
QY 121 GTPATKFKDLETLMMGKEKANIIVELGISVGTGQWEPKKTVTVAEAOFTAKTDKQFGT 180	
DB 150 GTPATKFKDLETLMMGKEKANIIVELGISVGTGQWEPKKTVTVAEAOFTAKTDKQFGT 209	
QY 181 EAAVESAKAVTTQKVEYKFAVEKLTKEDIKYTNKANNDVLVKEVTLSEDKRSATVE 240	
DB 210 EAAVESAKAVTTQKVEYKFAVEKLTKEDIKYTNKANNDVLVKEVTLSEDKRSATVE 269	
QY 241 LYSNLAKQTYTVNVNKGTEVAVGSLEAKTTEMAQDTVADEPTALQFTVKDENGTEV 300	
DB 270 LYSNLAKQTYTVNVNKGTEVAVGSLEAKTTEMAQDTVADEPTALQFTVKDENGTEV 329	
QY 301 VSPGIEFVTPAAEKINAKGEITLAKGSTTVKAVYKDKGVVAESKEVKVSAAGAAVAS 360	
DB 330 VSPGIEFVTPAAEKINAKGEITLAKGSTTVKAVYKDKGVVAESKEVKVSAAGAAVAS 389	
QY 361 ISNMTVAEONKADFTSKDFKONKNKVVYBGDNAYVOVELKDQFNAVTTGKVEYSINTEYAV 420	

```
|||||
Db 390 ISNMVVAEONKADFTSKDFKONNNKYEGSNAYVVELKQFNAVTTGKVEYESLNTFNAV 449
Oy 421 VDKATGKVTLSAGKAPYKVTVDKSKGALVSHVTEIEAFQAKMKDILEKTVALSTR 480
Db 450 VDKATGKVTLSAGKAPYKVTVDKSKGALVSHVTEIEAFQAKMKDILEKTVALSTR 509
Oy 481 DVTDLKVPKPVLDQYCKETAPVTYVLDKDKELKEOKLEAKYVRELVLNAGGEAEN 540
Db 510 DVTDLKVPKPVLDQYCKETAPVTYVLDKDKELKEOKLEAKYVRELVLNAGGEAEN 569
Oy 541 YVVLVLAKSGEKAKATLALTELKAPGAFSKFEYRGIDTELDRKYTEENQKNAATYVSLPV 600
Db 570 YVVLVLAKSGEKAKATLALTELKAPGAFSKFEYRGIDTELDRKYTEENQKNAATYVSLPV 629
Oy 601 DANGVLKCAEAAELKVTYTTNKEGKEVDATDAQVYVONNSVITVGGCAAGETKYTVVL 660
Db 630 DANGVLKCAEAAELKVTYTTNKEGKEVDATDAQVYVONNSVITVGGCAAGETKYTVVL 689
Oy 661 DDKLTITTHSFKYVDTPAKGLAVEFTSTLKEVAPNADLKALNILSVGVPATTTAKA 720
Db 690 DDKLTITTHSFKYVDTPAKGLAVEFTSTLKEVAPNADLKALNILSVGVPATTTAKA 749
Oy 721 TASNVEFVSADTNVVAENGTGKAGATSIYKKNLVKDKGKEQVEFDKAVQVAVSIKEA 780
Db 750 TVSNVEFVSADTNVVAENGTGKAGATSIYKKNLVKDKGKEQVEFDKAVQVAVSIKEA 809
Oy 781 KPATK 785
Db 810 KPATK 814
```

RESULT 2

S-layer protein precursor - Bacillus licheniformis
C:Species: Bacillus licheniformis
C:Date: 22-Oct-1996 #sequence_revision 01-Nov-1996 #text_change 15-Oct-1999
C:Accession: J04930
R:Zhu, X.; McVeigh, R.R.; Malachuk, P.; Ghosh, B.K.
Gene 173, 189-194, 1996
A:Title: The complete nucleotide sequence of the Bacillus licheniformis NM105 S-layer-en-
A:Reference number: J04930; M01D:97082965
A:Accession: J04930
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-874 <ZHU>
A:Cross-references: GB:U38842; NID:g1055336; PIDN:AAC44405.1; PID:g1055337
A:Experimental source: strain NM105
C:Comment: This protein is a glycoprotein. It functions as protective coats, molecular s-
C:Genetics:
A:Gene: slpA
C:Superfamily: S-layer repeat homology
F:1-29/Domain: signal sequence #status predicted <Sig>
F:30-874/Product: S-layer protein #status predicted <Mat>
F:157-209/Domain: S-layer repeat homology <SLR3>

Query Match

19.6%; Score 760; DB 2; Length 874;
Best local similarity 31.9%; Pred. No. 8, 6e-25;
Matches 268; Conservative 114; Mismatches 327; Indels 130; Gaps 34;

```
Oy 1 AGTFFPDVPAHMGIDISINLVKGAIVKNDKGFEGKELTRAEAATMAQIILNLPIDK 60
Db 30 AGKSPDPVPAHMAEDSINLVKGAIVKNDKGFEGKELTRAEAATMAQIILNLPIDK 89
Oy 61 DAKPSFADSGQWYTPPIAAVEKAGVTKGTG-NGFENGKIDRSMSLVEAKKLDTKV 119
Db 90 NAQPSFADANLWSSKXIYIAVEKAGVTKGTGKDNFYPEGKIDRSASFMSLVGAIVNLEKV 149
Oy 120 NGPAPATFKLETLNMCKEKANILVELGISVGTGDQWEPKKTIVKAAQAFIAKTDKQFG 179
Db 150 DGLTVTKFDDLRG-HMGEKANILVNLGISVGTGCKWEPKKSYSRAEAQFIATLTDKRY- 207
```

```
Oy 180 TEAAKVESAAVTTQKEVVKFSK-----AVEKTKEDIKVTNKANNDVLKVEVTLSED 233
Db 208 ---AKPENSDAVTNVAATPEPQLTLTGTLNKLTAEDVTL-----EGNKALALEA--SKD 258
Oy 234 KSNATVELYSNLAQOTYVVDNKKVKTETEVANGSLEAKTI-----EMADQYVADE 284
Db 259 GRSAAVTLSGKIAPNKELEPVKVK--GNFIVKVVYEVKKLRVEQLTFDDDRADQAAV--- 313
Oy 285 PTAQFTVDENGTE-----VVSPEGIEFY-----TPA-----AEKINAKGEITLAK 326
Db 314 -----FKLNDEKGNADIEVLIDIGHVYKFAVANNLDGTPANIFPGGTAESTGLANGIAE 368
Oy 327 GSTTIVKAVYKKDKQVVAESKEVYSAGAASISNMVVA-EQNKA-----DFTSKDF 379
Db 369 GKRYKVEQVTKRGITVSNMTGLEIKNLDAAEAFAIKDVFPAVDTRAGVYAKPLSGTDF 428
Oy 380 KQNNK-VYEGDNA---YVVELKDQFNAVTTGKVEKESLNTFVAVVDKATGKTVLSAGK 435
Db 429 TLNSKTLVAGEKAGIHKVVAQINKENKVVDPASISLKSNNP--GVLSVKNGETKAEPAAGS 486
Oy 436 APYKVTVKD-SKQKALVSHVTEIEAFQAKMKDILEKTVALSTRDVTDLKVPKPVLDQ 494
Db 487 ATLTVKGVDTKTFDFVVKTDTRKLTTVKANPD-----QLKVVDKRELPTYFTTDDQ 538
Oy 495 YGKEFTA-----PVT--VKVLD---KQCK-----ELKEQKLEAKYVRELVLNA 533
Db 539 YGDPFGANSGLAIKEVPQGTGVVVLVDTTNEGISIGTSSIKVKGENVAGCTIHQNP-NA 597
Oy 534 AGQEGNRYTVVLTKAKSGEKAKATLALTELKAPGAFSKFEYRG--LDTELDRKYTEENQKN 591
Db 598 SGEYGSLSHVEVTKSNGHAP---REL-----YSKAGKGGAADPTLIGAGNTVAYQLS 649
Oy 592 AMTVSVLPVDANGVLKGAEEAEKVTYTTNKEGKEVDATDAQVYVONNSVITVGGCAKAG 651
Db 650 NYTTEGVYADAADLAGEEFVGVNDKTIASAKIEGKTLKVTGKTAGV--TDVILTKRGATAG 707
Oy 652 ETKKVVYLDGKLTTHSFYVVDTPAPAKGLAVEFTST-----LKEVAPNND--LKAA 703
Db 708 HA-TITVTOENIQITSGFDVEE-----QFERKNYIDRVLDLVVSSDDQVNLGI 758
Oy 704 LNILSVGVPATTTAKATASNVEFVSADTNVVAENGTGKAGATSIYKKNLVKDKGKE 762
Db 759 KLVNI-STENKVRIVDESTBQCKVYLDNDNATDGNV-ALGVYTVAKNSDPTYSKEGND 815
```

RESULT 3

S-layer protein precursor - Bacillus stearothermophilus
C:Species: Bacillus stearothermophilus
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T14850
R:Jarosch, M.; Egelseer, E.M.; Maltanovich, D.; Sleytr, U.B.; Sara, M.
submitted to the EMBL Data Library, April 1999
A:Description: Nucleotide sequence of the coding region of sbSC, the S-layer gene fro
A:Reference number: Z18239
A:Accession: T14850
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-1099 <JAR>
A:Cross-references: EMBL:A055578; NID:g4581955; PID:g3025826; PIDN:AAC12757.1
A:Experimental source: ATCC 12980
C:Genetics:
A:Gene: sbSC

Query Match

11.5%; Score 445; DB 2; Length 1099;
Best local similarity 24.6%; Pred. No. 1, 8e-11;
Matches 223; Conservative 136; Mismatches 331; Indels 218; Gaps 41;

```
Oy 44 AEAATMAQIILN-----LPIDKDAKPSFADSGQWYTPPIAAV 81
Db 28 SQAATDVATVYSAQAQMKKAVYTYTSHVTFETGQFPDIKDVYAYAKRAK-QAYANAAVAV 86
```

QY 82 EKAAGVILKGTG-----NGFEPNKKIDRVASMASLLVEAYIKLDTKYNG-----TPPAT 125
| | | : : : : :
Db 87 NKAGGAKKDAVLADLAQIYYETTYVFKNPKSGEARVAT-YIDAYNVATKLDKMRQELKAAV 145
QY 126 KFPDLETLN-----WGKE-----KANLVDELGIS 149
| | | : : : : :
Db 146 DAKDLTKAEELHYHKISTELKTRIVYLDRVYGOSTELLRSTFKADAQAALDRLLIYDIYVA 205
QY 150 VGTGDOMEPRK--TVITYAANA-----OFIAKTDKOFGTE-----AAKVE 186
| : : : : :
Db 206 MKAREADQAVKACGLDKAKKALLDOVNOYYSKVTDAPFKRELQAKAADAKAYEALLPKVE 265
QY 187 SAKAVTTQKVEKFSKAVEKL--KEDIKVTINKANDKLYKEVTLSEDKRSATVELYSN 244
| : : : : :
Db 266 SVSAIDSTSKFYFTFKPVDKATAIPKNSITLTKGETKFLPKPSVSESGLTFATVLLYDT 325
QY 245 LAKQNTYTVOVNKGKTEEVANGSILEAKTIEMADOTVVADEPALOT----- 250
| : : : : :
Db 326 LVDGKTTYTTS--GLKDQTAGEFEFTSTNEFTYNRPV--PASITFNENKLPEDSAVDLT 380
QY 291 ---TVDENGTENVSPEGIEFVTPPAER-----INAKEITL-----AKGS-TVVKAV 335
| | | : : : : :
Db 381 KYTVVDAADANVIKS--GFELFETSSEKLTOGKFINTTGKRSVIYNAVYKGVNTGNVI 438
QY 336 YKDKGAVVASKEKVSAGAASISMTVAEQNKADFTSKDFKONKNVEGEDMAVVOV 395
| : : : : :
Db 439 LAVEDEKAAEVSSELKLRKNEKVEYTL---YANGNAFD-----KQGNISSG-TLTLTA 487
QY 396 ELKDQFNAAVTGGK----EYESLNTFEEVAVVDKATGKGYTVLSAGKAPVAYTVDSKGL 450
| : : : : :
Db 488 KFPDQYNELTGKAGAGDYTFESLNPEVLVV-APDGSVPIPVGVALVKKYGE----- 540
QY 451 VSHTVETIEAEQAOKAMKDIKLEKTNVALSTKDVYTDKVKRAPVLDQYCKEFTAPYTVKLDK 510
| : : : : :
Db 541 VTKITPTVYKANPVLKETIANDSTGVSAKGQKATRKV--TLADQYGNKFTGNVNV----- 593
QY 511 DGKELQEOKLEAKYVNBELVNLAAQOEAGANTV-----VLTAKSGEKEAKATAL 560
| : : : : :
Db 594 ----TSDKTETATVS----VSNSGIGOSEYTVTVNGVAEGSFITLIKSGTKVKV--VPV 642
QY 561 ELKAPCAFSPFVRGLDT-ELDKYVTEENQKNAMVSVLPVANGLVYLGAEAELKVT 619
| : : : : :
Db 643 NVYAGGVNAVYOILVLDLGKIDKSAATESPANNDVOLKYYAVDANGINV-GDINDVYITS 701
QY 620 --TNKECKEVDATAOYTVQNNSVYTVVGQAKAGETKYKTVVLDGKLITHSPKVVDPAP 677
| : : : : :
Db 702 EAIDTNCVINYNASSTANGDPIVYITDNGSKVKGE-TLTVLGLGYTLGTVDYEVIDT-- 738
QY 678 TAKGLAVEFTSLDK-EVAPNADLKAAALLNTLSV---DGPVATTAKATASME----- 726
| : : : : :
Db 759 TLKATVYTKKADLILEDADNGDALAKILLANDIDQNGNPWSDASATPNTNKLDQLKS 818
QY 727 ----FVSADPNVVAENGTV-GAKCATSIYVKMLVYVKDGKEQKV---BEDKAVOVAVSI 777
| : : : : :
Db 819 VLSGIYSSDTSVIGSVSNVDNMLKDDASI--SGLAVYKAGCTVTLTLEVNEBDSIAPAIVT 876
QY 778 KEAKPATK 785
| | | : : : : :
Db 877 K-APATIO 883

RESULT 4
A33856
surface-layer_125k protein precursor - Bacillus sphaericus
C:Species: Bacillus sphaericus
C:Date: 09-Mar-1990 #sequence_revision 09-Mar-1990 #text_change 15-Oct-1999
C:Accession: A33856
J.Bowditch, R.D.; Baumann, P.; Yousten, A.A.
J. Bacteriol. 171, 4178-4188, 1989
A:Title: Cloning and sequencing of the gene encoding a 125-kilodalton surface-layer prot
A:Reference number: A33856; MUID:89327128
A:Accession: A33856
A>Status: preliminary

Query Match	10.4%	Score 402.5	DB 2	Length 1176
Best Local Similarity	23.1%	Pred. No. 1.2e-09		
Matches 212	Conservative 139	Mismatches 374	Indels 191	Gaps 36
Qy	16	DSINIVYVEGAVKNDGKMFEEPGKELTREAATMAQIILNPIDDKAPSFADSGQ	WY 74	
Db	45	EAVOSLVYDAGVYIQGGANGNFPLKITSRAEAATFTFMALTEAEEDV--NFADVADAMV	102	
Qy	75	PFPIAAVEAGYIKG--TGNGFEPPNCKIDVSNASLSIVYKIDTGVNGTPAKFPDLETL	133	
Db	103	YPAIAATVYENGIFEGVSATFEAPNKOLTRSEAKILLVDAFELEGECD--LSEFADASTV	159	
Qy	134	N-WGKEKANILVELGISVGTG-----	DQME 157	
Db	160	KWMASTYELIAVANGYIKSEANGKTNLPNAPITRQDFAVVFSRTIENVDTAPKVDKIE	219	
Qy	158	-----PKTYTKAAEAQFIKTDKQCG-----TEAAVESAK	189	
Db	220	VVDAKTLNLTSLDGRKETVYLTLEKALEPENKETEVEFKIKDVEKAKVTVYVVTATAVKSVS	279	
Qy	190	AVTTQOYVEKFSKAVEXLKLKEDIKVTNNKANNCKVLKYEVTLSSEDKRSATVELYSMLAAKQ	249	
Db	280	ANILEVEVVEFPGYVDEKTEAED--AANYALKSGKTIKTSYLAADKNTATVITLTKLNNK	337	
Qy	250	TYTVADN--KVKCTEVAAGSL-----FAKTEMAQQTVADEP---TALQF	290	
Db	338	ADAISISNNKADDKELNVKNNVETFAVDNKIPEVTEKSLGTAAVYTLSEPPENLSSTNF	397	
Qy	291	TYKDE-----NGTEVVSPEGIEFVTPAEXIKNAKGEITLAKGSTVIVKAVYKQD	339	
Db	398	TLDGRKAYFGNNVMGAGNKTVILTPYSSALSVDGDKLTVSGAKDFAGFVSLSTHEFK--	455	
Qy	340	GRVVAESKEVKVSAEBAVAASISNNITVAQNNAD-----FTSKDQFONKKYE--CD	389	
Db	456	---VEDDEAPPTVTEATALETTLVTLTFESDDIMDTIVKASNNYWKSGDSKKESEFERIAD	512	
Qy	390	NAVVOVELDQFNAVTTGKVEY-----ESLNTTEVAVVDKATGKVTLSAGCAPKVTWK	443	
Db	513	NKTKYF--FGSEKTLPTGTGVYVVEDIKQYSDNKKIAKDKVTYVTEPIDQTRFVKKVTAL	571	
Qy	444	DSKG--KALVSHTVETIEAPAKAMKDIKELTNAVALSTKQVDTLK--VKAPVLDOYSGKEPTA	501	
Db	572	DEKTIIVTFSTKIVDGESAIKTGNITVKKDKDQVVSVDKTYVDSKSDSKSVIILDKSVSGV	631	
Qy	502	PATYVAVLD--KDGKELKEQKLEAKYVNRVELVLAAGQEAQNTVYVLTAKSGKEAKATLAL	560	
Db	632	ENTITITKNVKDKATKLNNTMLD--YTGK--FTRSDKEGDPYHVINADA--TAKKVVLFK	684	
Qy	561	ELKAPGA---FSKFEVRGLDT--ELDKYVTEENOKNANTVSVLP-----VDANG	604	
Db	685	DKKMDAASLADYSNYLYKINDTLQITSEVALILSNSNDATVYTIIFAEITIKDDVYFASG	744	
Qy	605	LVLKGA---EAAELKY---TTTNKEGKEVADTAQVTVONNSVITVGGAAAGETY--K	655	
Db	745	KAISGSGKVNVELQYMGVKDTSIGNVHKFKFGENSENKITLSTPS--TPRLKLAIDKDYDAK	802	
Qy	656	VIV--VLDDGLITTHSKVVDIAPTAAGLAVETYSISK-----EVAP	696	
Db	803	YTAELDRKTVKVFSTVINSA---AANATSEBKHIDSIOVNGSTVTVYKFKDEINT	857	
Qy	697	NA---DLKAAALNIIISVDQVPAT--TAKATAVNEVESADTVVAENGTGAKGATSIYV	751	
Db	858	NASDDDLKNLKSLVDIAGNESTNNTPIAKGAINLLDSVAAPVYVVGEPVDDKEITITFTFS	917	
Qy	752	KNLTVVKGDEQKQVEF	767	

Db 918 ENLTSVSGEVLSTDF 933

RESULT 5

F59097

hypothesized protein px01-54 - Bacillus anthracis virulence plasmid px01

C:Species: Bacillus anthracis

C:Date: 12-Nov-1999 #sequence_revision 12-Nov-1999 #text_change 11-May-2000

C:Accession: F59097

R:Okimaka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koehler

J.; Bacteriol. 181, 6509-6515, 1999

A:Title: Sequence and organization of px01, the large Bacillus anthracis plasmid harbored

A:Reference number: A59091; MUID:99445483

A:Accession: F59097

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-404 <OK>

A:Cross-references: GB:AF065404; NID:94894216; PIDN:ADJ2358.1; PID:94894270

A:Experimental source: strain Sterne

A:Note: similar to S-layer precursor, surface layer protein (814 aa), B. anthracis (p490

C:Genetics:

A:Gene: px01-54

A:Genome: Plasmid

Query Match 7.2%; Score 279.5; DB 2; Length 404;

Best Local Similarity 25.8%; Pred. No. 4.2e-05;

Matches 107; Conservative 68; Mismatches 161; Indels 79; Gaps 18;

QY 5 FPDVADHWGIDSYNYLVEKGAVKNDKGFEPGKELTRAEATMMAOILNLPDKDA-- 62

Db 48 FNDVADHWGIDSYNYLVEKGAVKNDKGFEPGKELTRAEATMMAOILNLPDKDA-- 106

QY 63 KSFADSOQWITPPIAAVEKAGVIGKGTNG--FEPNGKIDRVSMASLVEAVKLDTRKVG 121

Db 107 KMFPSIKHMEKEKELTALAKBGLVAGYGBGFGPDILITREMAQVILNNAEF--KA 162

QY 122 TATATKEDLETINMCKEKANILVELGISVGTGDQ--WEPRKYTKAEAAOPIAKTDKQFGT 180

Db 163 TTTSTSTIDKNSMALKAISALEENGVTITGTGDKLSPYAHVTRQYSOLFENS----- 216

QY 181 EAAKVESAKAVTQAKVEKFSKAVEKLTEDIKVTNKKAN--NDKVLVKEVTTSED--KRS 236

Db 217 ---INVLKERTKPEVKPEKPEKPEKPEKPEKPEKPEKPEKPEKPEKPEKPEKPEK 272

QY 237 ATVELSNLAOKOTYVDVKKVGTETVAAGSLKAKTIEMADQTVVADEPTALOFTVKDEN 296

Db 273 KYINSSSSIOA--OSLINEVKKKNTNLKAAEV--GVIRLVD-----N 312

QY 297 GREVVSPEGI---EFTVPAAE--KI---NAKGETTLAKGTSTVVKAVYKKDKGV--V 343

Db 313 G--MYLPEGLAKQFYIVPINEGSFKYIFLDNNDATVELAKKWTTLKSDLNDEELOGT 370

QY 344 AESEKVEKVSAGAAVAVASISMTVAEONKADFTSKDFQNNKVEEGDMAYVOVELK 398

Db 371 VESHKLNTFEKGNVYKVRIGNST-----ADHM-----MYOVESK 404

RESULT 6

T30311

S-layer protein - Clostridium thermocellum

C:Species: Clostridium thermocellum

C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999

C:Accession: T30311

R:Lemaitre, M.; Miras, I.; Gounon, P.; Beguin, P.

Microbiology 144, 211-217, 1998

A:Title: Identification of a region responsible for binding to the cell wall within the

A:Reference number: Z20818; MUID:98129094

A:Accession: T30311

A:Status: Preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1036 <LEN>

A:Cross-references: EMBL:U79117; NID:93493463; PID:93493464; PIDN:AAC3404.1

C:Genetics:
A:Gene: slpA

Query Match 6.9%; Score 268; DB 2; Length 1036;

Best Local Similarity 20.5%; Pred. No. 0.00044;

Matches 217; Conservative 133; Mismatches 325; Indels 392; Gaps 50;

QY 16 DSINY-----LVEKGAVKNDKGFEPGKE--LTRAEATMMAOI-----LNLPIDK 60

Db 27 DSFSEKKEKELIYRGVIGKGTSETVEYNLBECKLDKQGVVLLRLFGEDDALEIPMDE 86

QY 61 DAK---PSFASQ--GQWTPPIA-AVEKAGYIKGTNG--FEPNGKIDRVSMASLVE-- 111

Db 87 AAQTLAAKFKDADADIADMAQROVAAVAVK--GYVKGYPDGTFLPMDLNLGACSLILOOL 145

QY 112 -----AYKLDT--KVNGTPATRKDLETIN-----MGKEKA 140

Db 146 GYDGFVDEAAVYKLOEFGGLTAEQAFAFNKNGINRDSMGVIAFSAQAVYKATGKTVI 205

QY 141 NILVELG-----ISVGTDQWEPKRTVTKAEAAOPIAKTDKQFGTEAKVESAKAV-- 191

Db 206 EVLVENGVSKELEIETELGVLLKAIKEVKALDAVKV-----OVGKEPVLPVEVEVYEDD 259

QY 192 TTQKVEVK-----SKAVEK--LTKEDIKVTN--KANND 221

Db 260 TTEKLAVEPVTVDTSVGEQIEGTIKGASGLAYREPRTATKYIYTPPELOVDVAKAPVL 319

QY 222 KYLVKE-----VTLSEDKRSATVELYSNLAOKOTYVDVVK 257

Db 320 KELVIEFNGEAVSKADEKSSYSVEDNTIELVTVYSEDKTIVTLVAGAMAAEEIEVTI-- 377

QY 258 VKTEVAVGSLEAKTITEMADQTVVADEPTAL--QFTVK-----DENGTE 299

Db 378 --KTATGLKEEYKTVVPADYENPEAESIALIGPNSFEIKFSEPVOSSDAEVLVNDGYI 435

QY 300 VVSPE-----GIEFTVPAEKINAKG-----EITLAKGS--TT 331

Db 436 YVSEELSDQYRTLYVELGVSSINEGTYYKVGVDAGNIMRTFTFDEYVKDTPPT 495

QY 332 VKAVYRKDKGVAAESKEVKV---SAEGAAVA-----SISNVTVAEONKADFTSKDFQ 381

Db 496 AKVKEATQNKVVIIEFNEPATRDGYSGEDEALTFDYFYQYSSKPKTVVAASD----- 547

QY 382 NNKVV-----EGDAAVYVELKDQFNATVTGKVEESLNTAVAVDKATGKVTYLSAG 434

Db 548 NNKVVYLLYSEDDNDGY-----FVILLPVG 573

QY 435 KAPVYTVKDSKGRKALVSHVEIEAPAKAMKDIKLEKTVALSTKDVDTLKYKAPVLDQ 494

Db 574 NVITIT-LKEVDDAVY-----DAMGNKLESPLKILAT--VAADNEAPTVKSVTAEAEDK 625

QY 495 YGKEFTAPVTK-----VLDKQKELKEQKLEAKYVRELVLNNAOGDAQNTYVLT 546

Db 626 IIVVYSEDEVNENAKDKDNVYIKKDKED-----TAISS 660

QY 547 AKSGEKEKATTLAELKAPGAFSFEYRGV--DTELD---KYVT---EENOKNMTVSVL 598

Db 661 ITYDSNETKVTIVDEKLSGKTTIDIKIKDTSVSENEKAVYIEEYVDKTAFTIEEV 720

QY 599 PVDANGVLKGAFAELK--VTTNKEG--KEVDATDAQVTV-----QNNSVITVQ 646

Db 721 TFDVNTIYVYSEAMSKNGSVLKNKNYKLVDDNDKVEIKKIELGSDKNNKRVIVDS 780

QY 647 GAKAGETVYKTV-----VLDCKLTTSF-----KYV 673

Db 781 DVDNLVYVELTIANVEDEAGNAISAPDVKAKKLSDEQAPESIEIRLISKEIEIYINKIL 840

QY 674 DTAPFAK-----GLAVEFTSTSLKVAIPA-----DLKALNLITL 708

Db 841 DKATVETKIDFEVERGSKNVALTRISSITTYDGKTIYGVLPDANRPANSODITGTYTLYI- 899

QY 709 SVDDVPATTAKE-----ATASNVFVSADTNVVAENGTVG---AKGANSTYVKNL----- 754

Db 900 -VGEIKSDTGKATGAVSPVDKFPSPFVSANGVYGASAKKGFLLFDEDIKELNNS 958
OY 755 -----TVVKDKG---EQKVEFDKAV---QVAVSIX 778
Db 959 AGIGATDLVYKNSKTLLEGIDYDAIDNKITVITLK 995
RESULT 7
A56143
surface-array protein homolog sapA2 - Campylobacter fetus (strain 82-40 LP3)
C:Species: Campylobacter fetus
C:Date: 03-Nov-1995 #sequence_revision 03-Nov-1995 #text_change 08-Oct-1999
C:Accession: A56143
R:Dworkin, J.; Tummuru, M.K.R.; Blaser, M.J.
J. Bacteriol. 177, 1734-1741, 1995
A:Title: A lipopolysaccharide-binding domain of the Campylobacter fetus S-layer protein
A:Reference number: A56143; MUID:95204338
A:Accession: A56143
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1109 <DMO>
A:Cross-references: GB:S76860; NID:9913763; PIDN:AA833871.1; PID:9913764
A:Note: this gene appeared to be silent in strain 82-40 LP3
C:Genetics:
A:Gene: sapA2

Query Match 6.7%; Score 259; DB 2; Length 1109;
Best Local Similarity 21.2%; Pred. No. 0.0012;
Matches 202; Conservative 122; Mismatches 316; Indels 312; Gaps 46;
OY 11 DHMGDSINYLVEKGAVKGNCKMFEPEKELTRAEATMAQIINLPIDK----- 60
Db 191 DEAGLTKALTTENDTITTEGGDLISGVYGTAESEF-----LN-PGDKIDGAGNDVL 243
OY 61 --DAKSPADSGQWTPFLAVERKAGVIGK--TGNGFEPNGKIDRVSNASLVE----- 111
Db 244 KYDLKNNFGLKDDGY---IKNIEKLSLTVSSVSNRFTPAKG-IDGIQVVALSGEKISV 299
OY 112 ---AYKLDTRKVNCTPATKFK----- 132
Db 300 TNLANTVDVENVNGFKGTGNFVDSITADKVLDSADYONLKVNGVGAKGASVATADKIET 359
OY 133 LNN-----GKEKANILVE--LGISVGTGDQWEPEKKTVTAKAAQFIKTDKQGTGTE 181
Db 360 LNLNTGSGSFVSADVASISVKGANLISLATG-----AKTTT-LDASSFGALDADLSTS 413
OY 182 AAKVESAKAVTQKVEVKSFAVEKLEKEDIKYT-----KKAANDKVLV----- 225
Db 414 A-----SVTS---IKGGNGNDKITIKDVAVNAVLAIDGAGNDELVIKGSTADTLOPTL 462
OY 226 -----KEVTLSEDKRSATVEL-VSNLAAR-----QTVTVVNVKVGKTEVA 264
Db 463 TNEKVTIDGNKTDLTLSLKAQSVTELSFKNIKATVYTESNGVVEYNIILANNATDKAVT 522
OY 265 VGSLEAKTIEMAD-----QTVVADEPTALQF-----TVKDENGTEV- 300
Db 523 INDESLKTIINFSVDKGSVAAGKIVADKATELFTINSKVTLASDPAVVAQANATKIDI 582
OY 301 -----VSPGIEFVTTPAAEKINAKGETTLAKGSTTYTKAVYKKDGKVAASEKVKVS 352
Db 583 NNAKQDTVGLTGGVAKLTDLT--VNNKGAPALTGANATDLDV-----KNLSVN 629
OY 353 AEGA-AVASIS-----NMTVAEQNKADP-----TSKDFQONNKVYEG 388
Db 630 TBEAFSIAIATSLKLNINLSNGVSADLNSVNGTATLASLEANINVSGEFKICTTAKG 689
OY 389 DNAYVOVELK-DQFNAVTTGKVEYESLNTFVAVVDRKATGKVTY-----LSAGKA 436
Db 690 D-----VDFEIEHWGALTICAITSSGTNASV-IISATGNVTLGAVSAGQNLTLNAGNF 743
OY 437 PKVYTVKDKSGKRL-----VSHTVLEAFQAQAMKDIKLEKTNVALSTVDVYDLKAKP 490

Db 744 LGNITIGALGDIVSDLVGGLCTINSDA-----NNKVSITSNEVT----- 784
OY 491 VLDQYGEFPAPYTVKVLDDGKELEOKLEAKYVNEVLVNAAGPAGVYFVLTPAKSG 550
Db 785 ---YVSEISKNVVEITAAAGGTDLNAQVIGGAAADALITIGKD-----DQITTA-SG 835
OY 551 EKEKATLLELKPAGFSFEVRGID-----TELDKYVTEBNKNMNTVSLVPDA- 602
Db 836 DLSG-GTLTLTTEATKLSLSDISGVKGTGNVAIELKRAV-----QGNKTDVSVQSGDAA 890
OY 603 -----NGVLKGAEEAELKVTYTNKEGKEVDATDAQV-----V 636
Db 891 EQITVTSAAISLTDIKISGDLGAGANTVTPDPAADLKTDISLSATGCTLASTITLV 950
OY 637 QNNSVITVGGAKAGETTYKTVVLVDGKLITTHSPKVVDTPTAKGLAVEFTSYLSKEVAP 936
Db 951 NANTAITSVKSGISGADI--ITVVSANKAVAILDGK-DTVID-----KVDYSIKISPSKN 1002
OY 697 MADLKAALL--NILSVGY--PATTAKATASNEFVSADTVVAENGTVG 742
Db 1003 DASIKADLVISITNALSGDQIVLKGATSIKDRGD-----LSEGANILALGKLG 1050

RESULT 8

A34355
cell surface protein precursor - Acetogenium kivui
C:Species: Acetogenium kivui
C:Date: 15-Jun-1990 #sequence_revision 15-Jun-1990 #text_change 08-Oct-1999
C:Accession: A34355
R:Peterson, J.; Peters, M.; Lottspeich, F.; Baumeister, W.
J. Bacteriol. 171, 6307-6315, 1989
A:Title: S-layer protein gene of Acetogenium kivui: cloning and expression in Escheri
A:Reference number: A34355; MUID:90036724
A:Accession: A34355
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-762 <PEP>
A:Cross-references: GB:M31069; NID:q141844; PIDN:AAA21930.1; PID:q141845
C:Superfamily: S-layer repeat homology
F:32-85/Domain: S-layer repeat homology <SLR1>
F:96-150/Domain: S-layer repeat homology <SLR2>

Query Match 6.6%; Score 255.5; DB 2; Length 762;
Best Local Similarity 22.7%; Pred. No. 0.00099;
Matches 177; Conservative 89; Mismatches 275; Indels 239; Gaps 35;
OY 5 FPDVPADHMGDSINYLVEKGAVKGNCKMFEPEKELTRAEATMAQIINLPIDKRAKP 64
Db 96 FKDVPOHMAVVGQINLAVYKGLAQGVNGKFPDPSLELYAOLAFLVALGF--KDL-- 150
OY 65 SFPSDQGWYTPFLAVERKAGVIGKTNNGFEPNGKIDRVSNASLVPAYKLDKFNCTPA 124
Db 151 -----DWPYGLAKQDDGLVHGLNLA--NCVYKRGDLALILDBALEV-----PM 194
OY 125 TKFKDLETLNMGREKANILVELGISVGTGDQWEPEKKTVTAKAAQFIKTDKQGTBAK 184
Db 195 VKYVD-----GKE-----VLGEPLIS-----KVATKAERY-VIATNAQDSVBEKG 234
OY 185 V-----ESAKAVTQKVEVKSFAVEKLEKEDIKVTNKAANDKVLKRE-----VTLSDKR 235
Db 235 VAVLDKDGKLTITINAGLVDFSEYLGK---KVITYSERFDPVYVAAGNDVVSFTGQD 290
OY 236 SATVELDSNLAQOTYTVVDN-----KVGK-----TEYAV-----GSLFA 270
Db 291 SVGTTVYKNDKNTALVVDNNAVLYNGYLTAKVSKVYVREGAEVTTINNNTLYNGSYDN 350
OY 271 KTIEMADQTVVADEPTALQFVVDENGTEVVSPEGIEFVTTPAAEKINAKGEITLAKGTS- 329
Db 351 STI-----VYNDVQSGDKYLRNDSN-----YELKGTVYVTVGAVSK 385
OY 330 -TVYKA-----VYKKDKGV-----AESKEVYVSGAAGVASISMWTV 366

```

Db      386 VTDIRKANDYIYGYQYDVNNGVTVIYVNRQVGTGTEKSYSGSTYKASIDNV- YTV 444
      367 AEONKAD-----FTSKD-----FKONKNVYEGDNAYVOVEL 397
      445 ADNNVWNOLEPGKVVYILNKDNYVIGISSTTTTAVNVAIFREKSDPPTAMFAKVLIL 504
      398 KDO---FNAVTTGKVEYESLN---TEVAVDOKATGRVYL-----SAGKAPVK 439
      505 PDAAEKVFADAVSD--YYDRVNLAEGTIVTYVDANGKLANDIORANDOPSSASAYKADAK 562
      440 VYAKDSGKALVSHVTEIE--AFQAKMKDKIKLEKTNVALSTKDVDTLKAAPVLDQY-- 495
      563 VLREGSTTYITDNTVTLNNTSDGFKALKLTDL-----KATNLNVAKI--VADNNV 612
      496 -----GKEETAFVYKVLIDKDKGKLEOKLEAKYVNRRELVLNAAGOEAGNYTVLTA 548
      613 AKYVVFNNASVSTSTTVY-----AYTGTADVYVNGSTFTRLTVLENGQ 658
      549 GGEKAKATLALELKAPGAFSKFEVRGLDTLDDKYVTEENKNAMTVSVLEVDANGVLK 608
      659 TKTYDANAOLATNY-----THKAVVLTLTNKNIANI--ALPTVASGVKLT 701
      609 GAFAELKVTITNKEGKVD---ATDAQVTYQNNSVITVGOGAK---AGETKYTVV 659
      702 NIDQANLRTTITNKGKYLDPNFIVVDJNGNKGSLDITKDTGVVLTNDGVKVFVIEIV 761

RESULT  9
140468 surface layer protein sbas - Bacillus stearothermophilus
C:Species: Bacillus stearothermophilus
C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 15-Oct-1999
C:Accession: 140468; S34365
R:Kuen, B.; Sleytr, O.B.; Lubitz, W.
Gene 145, 115-120, 1994
A:Title: Sequence analysis of the sbas gene encoding the 130-kDa surface-layer protein
A:Reference number: 140468; MUID:94320770
A:Accession: 140468
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1228 <RES>
A:Cross-references: EMBL:X71092; NID:g312729; PIDN:CAAS0409.1; PID:g312730
C:Genetics:
A:Gene: sbas

Query Match      6.4%; Score 248.5; DB 2; Length 1228;
Best Local Similarity 22.4%; Pred. No. 0.0036;
Matches 201; Conservative 109; Mismatches 309; Indels 279; Gaps 44;

```

```

Db      610 EFKTQDLTA--PTVIVTSKNGDGLKVT--EAOEFTVKESENINTEFNATTVSGSTITYG 665
      331 -----TVKAVYKKDG-----KVASEKE----- 348
      666 QVAVVYKAGANISALTASDIIPASVEAYTGODGTYKVVANNOLEPNNGYKLVVGKGATA 725
      349 -VKVSAGAAVASISMTVAEONKADFTSKDFKONKNVYEGDNAYVOVELKDFGNATYG 407
      726 PVKDAANANLTATNYIYTFTEGQ-DVTAPlV---TVFKGDS-----LKDA-DAVTT- 773
      408 KVEIESLNTVAVVDKATGKTVLSACKAPVYV---KDSKGLVSHVTEIEFA 461
      774 -----LTNVDAGQ--KFTIQSSEELKTSIG-SLVGKXTVTEKLT 809
      462 OKAKMDIKLEKTNVALSTKDVTDLKVPAPVLDQYKGEFTAPVYKVLIDKDKGKLEOKLE 521
      810 NNGWVD--ACTGTTVS-----APKPDANGKVAAVTLGLDNNDKAK-LRLV 856
      522 AKYVNRRELVLNAAGQ-----EAGNTYVLTAKSGEKAKATLALELKAPGAFSK 570
      857 VDKSSTDGIADVAGVNIKEKDILIRYNSWRHTVSVKAAADKQDQ-----NSAAPP- 908
      571 FEVRGLDTELD--KYVTEENKNAMTVSVLEVDANGVLKGAFAELKVTITNKEGK 627
      909 -----TSTAIDTYSLVEFNE-----TDIAEVKRENIVVDAAGNAVAGVTALDGS-- 956
      628 DATDAQVTYQNNSVITVGOGAKAGETKYTVVLDG-----KLITTHSFKVVDPAP 677
      957 -----TNKFVPTPQELKAGTVYSVTI--DGVVRDVGNTISKYIT--SEKTVSANP 1003
      678 TAKGLAVEFTSLEKVPANMDLKAALLNLSVDPGP-ATTKATVASVNEEVSAD-TNVV 735
      1004 T-----LSSISIADGANNVD-RSKTITIEPSDVPNPPTITLKKADGTSFNITLVN 1055
      736 AENGTVGAGAT--SIYVKNLTVVKDG-----EOKVEPDKAVOVAVSIEKAPKA 783
      1056 NENKTYKIVFHKGVLDEFTQYELAVSKQFGCTDIDSKVTP---ITGSVATDEVKPA 1110

RESULT  10
S33441
EP protein - Streptococcus suis
C:Species: Streptococcus suis
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 15-Oct-1999
C:Accession: S33441
R:Smith, H.E.; Reek, F.H.; Vecht, U.; Gielkens, A.L.J.; Smits, M.A.
submitted to the EMBL Data Library, May 1993
A:Description: Repeats in an extracellular protein of wek-pathogenic strains are abse
A:Reference number: S33441
A:Accession: S33441
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1822 <SMT>
A:Cross-references: EMBL:X71880; NID:g298031; PIDN:CAAS0714.1; PID:g298032

Query Match      6.3%; Score 245.5; DB 2; Length 1822;
Best Local Similarity 21.2%; Pred. No. 0.0082;
Matches 201; Conservative 120; Mismatches 345; Indels 281; Gaps 41;

```



```
Db 1084 ESANKVAEEAAKATTAIDKASPDVAQVEDGVAAINLITAKAKAGVIAAKLADEIK 1143
OY 200 -----SKAVE--KLTFEDIKYNKANNDKVLVEVTLSEDKRSATVEL-----242
Db 1144 KLEDDKAEAEKALIDASTMTTEKAIAKKALQDVDDGKALEDAARATVEIHEATTTEK 1203
OY 243 ---SNLAQOTYVDVNVKGTVEAVSGLEAKTIEMADQTVVADEPTALQFTVK-----293
Db 1204 AKAAELAGEKSLT-DTGKEARDAVEL-----AKDKELAKEAIPTREEEPAFIVEKLAEDTR 1258
OY 294 -----DENG-----TEVSP-----EGIEFTVPAAEKINAKETITLAKST 330
Db 1259 KALEDPNLSDEDKQAEIKKLTDAVAKTLATIRDMADKROEAEKQALADLEAKETQK 1318
OY 331 TVK-----AVYKKDKGVAAESKEVY-----VSAEGAIVASISNMTVAEONKADFT--375
Db 1319 IADKAAIDRLTILVKGELATKQDAKNKIAKDAAAAKEIASNPILTDAE--KTTTDDA 1376
OY 376 --SKDEKQNN-----KYEGDNAYVOVELKDQFNAV--TTGKVEYESLNTTEVAV-420
Db 1377 VDAEVAKKANDAIASATSPADVOKEEDGVAIAIEDVLDAAKODAKNKIAKDAAAAKGAIG 1436
OY 421 -----YDKATGKVT-VLSGKAPVKYTVVDSKGLVSHTEIEAEPAQK 463
Db 1437 SNEPLTDAEKKITFTDAVDAEVAKANDAIASATSPADVOKEEDGVAIAEDV-LDAAKOD 1495
OY 464 AMKDIKLEKTNVALSTKDYDLKVKAPVLDOYGEFT-----APTVVLDKDGKELK 516
Db 1496 AKKKIKKE--SDAKSAIDAN-----PNLTDAREKESAKKAVADAKATAIDASTSPVE 1548
OY 517 EOKLEAKYVN--RELVLNAAQEGAGNYTVVLAKSGEKEAKATLAEKAPAFSKEFVR 574
Db 1549 AOSAERKGVGSIADQVLDKAKODAKN-----KIAKEVAAAKEAIDAN-----NLS 1594
OY 575 GIDTELKDYTEENOK--NAMTVSVLPVANGVLKGA-----EAEELKVTTNKE 623
Db 1595 DAEKESKRAVNDADAKTTDAIDASTSPVEAOSAEDKGVGSIRODVLDAA--KODAKNKI 1652
OY 624 GKEVDATDAQVTVONNSVITVGOGAKAGETKYVTVLDGLITTHSEKVVDTAFTAKGLA 683
Db 1653 AKESDAKSAIDANPNLTDAREKESAK-----KAND-----1682
OY 684 VERTSISLEKVAENADIKALNILSVDPATTAKTASNVE-----FVSAD-----731
Db 1683 -----ADAKAA-----TDAIDASTSPVEAOSAEDKGVGALAKDILDAKQ 1722
OY 732 --TNVAENGTGAKGATSIYKNLTVVXKDKKE--QKVEFPKAVOVAV 775
Db 1723 DAKNKIAKE-----AESAKSVIDSNPILTDAAKEAKSEIDKAVEEAI 1765
```

RESULT 11
T30296
R27-2 protein - Trypanosoma cruzi

C:Species: Trypanosoma cruzi
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Mar-2000
C:Accession: T30296
R:Otsu, K.; Donelson, J.E.; Kirchhoff, L.V.
Mol. Biochem. Parasitol. 57, 317-330, 1993
A:Title: Interruption of a Trypanosoma cruzi gene encoding a protein containing 14-amin
A:Reference number: Z20813; M0ID:93185082
A:Accession: T30296
A:status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1128 <OTS>
A:Cross-references: EMBL:L04603; NID:g385171; PID:g1256742; PIDN:AAA96494.1
C:Superfamily: neurofilament triplet H protein

Query Match 6.2%; Score 241; DB 2; Length 1128;
Best Local Similarity 25.5%; Pred. No. 0.0067;
Matches 191; Conservative 90; Mismatches 353; Indels 116; Gaps 34;

```
OY 99 KIDRVSNASLIVE---AYKLDTKVNGTPATKREDETLNMGKEKANIIVELGISVGTGD 154
Db 15 KLDIAINATLTLEBRNQAYAAHKAEEKAKTQRLITFE--SENIMUKKRPNDVAENRD 72
OY 155 QMEPKTV-----TKAAQOFIATDKQFGTEAAV-----ESKAVTTQKVE 197
Db 73 KKNSETAKTDEVEKORAAARAV--ETEKORAAEAATKVAEAKRAAEAKAETETEKOR 131
OY 198 VKFSKAVEKLTKEKIDKVTNKA--NNDKVLVKEVT--LSEDKRSATVELSNLAQOTYVDV 255
Db 132 AAETKVAEEKOKAAEAATKVAEAEKORAAEATKVAEAEKORAAEAKKVAEAEKORAAEA 191
OY 256 NKVGTE---VAVSGLEAKTIEMADQTVVADEPTALQFTVDENGTVEVSPGIEFTV 310
Db 192 TKVAEAEKOKAAEATKVAEAEKOKAAEATKVAEAEKQ-----KAAEATKVAEAE-----KQ 242
OY 311 PAAEKIN-ANGELITLAKGTITKAVYKK--DGKVAEESKEVKSAGAAVASISNMTVA 367
Db 243 KAAEATKVAEAEKOKAAEATKVAEAEKOKAAEATKVAEAEKOK--AAEATKVAEAEKOKAA 301
OY 368 EONK-ADFTSKDEKONKRYEGD-----NAVVGVELKDQFNAVTTGKVEYESLNT--TE 417
Db 302 EATKVAEAEKOKAAEATKVAEAEKOKAAEATKVAEAEKOKAAEATKVAEAEKOKAAEATK 361
OY 418 VAVVDKATG---KYTVLSAGKAPVKTVKDSKGLVSHTEIEAPQAKMDIKL---470
Db 362 VAEAEKOKAAEATKVAEAEKOKAAEATKVAEAEKOKAAEATKVAEAEKOKAAEATKVAEA 421
OY 471 EKTINVALSTDYDLKVKAPVLDOYG--KEFTAPVTVYLDKNGKRELKQKLEAKVNR 527
Db 422 EKOKAAEATKVAEAEKOKAAEATKVAEAEKOKAAEAT--KVAEAEK--KQAAEA-----471
OY 528 ELVLNAAQEGAGNYTVVLFAKSGEKEAKATLAEKAPAFSKEFVRGLD---TELDKY 583
Db 472 TKVAEAEKOKAAEATKVAEAEK--QKAAEATKVAEAEKOKAAEATKVAEAEKOKAAEATK 530
OY 584 VTEENOKNMTVSVLPVDANGVLKGAELKVTTNKSEKVEDATDAQVTVONNSVIT 643
Db 531 AEAEKOKRAEATKVAEAEKQ---KAAEAT--KVAEAEKQ--KAAEATK-----VA 573
OY 644 VGGAKAGETKYVTVLDGLITTHSEKVVDTAFTAKGLAVEFTSLUKEVAPADLKAA 703
Db 574 EAEKOKAAEATKVAEAEKQ-----AAEATKVAEAEKOKAAEAT-----KYAEAEKOKAA 623
OY 704 -LNLISVDGPATTAKTASNVEFVSADTNVVAENGTGAKGATSI-----YVKNLTV 756
Db 624 EATKVAEAEKOKAAEATKVAEAEKOKAAEATKVAEAEKOKAAEATKVAEAEKOKAAEATK 683
OY 757 VKDGKEQVEEDKAVQVAVSIKE-AKPAIK 785
Db 684 VAEAEKOKAA--EATKVAEAEKOKAAEATK 711
```

RESULT 12
T34418

hypothetical protein F12F3.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34418
R:Patton, B.; Wohlmann, P.
Submitted to the EMBL Data Library, July 1998
A:Description: The sequence of C. elegans cosmid F12F3.
A:Reference number: Z21521
A:Accession: T34418
A:status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3488 <FUL>
A:Cross-references: EMBL:U80022; PIDN:AAC25885.1; GSPDB:GN00023; CESP:F12F3.3
A:Experimental source: strain Bristol N2; clone F12F3
C:Genetics:
A:Gene: CESP:F12F3.3
A:Map position: 5

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2002, 07:52:42 ; Search time 66.74 Seconds

(without alignments)
431.254 Million cell updates/sec

Title: US-09-754-947-1

Perfect score: 3885

Sequence: 1 AGKTFPPVADHMGIDISIN.....EFDAKOVAVSIRAKKPRATK 785

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3847	99.0	814	1	SLAP_BACAN
2	760	19.6	874	1	SLAP_BACLI
3	402.5	10.4	1176	1	SLAP_BACSH
4	255.5	6.6	762	1	SLAP_ACEKI
5	248.5	6.4	1228	1	SLAP_BACST
6	234.5	6.0	1861	1	APU_THEUT
7	232.5	6.0	1087	1	XYNX_CLOTM
8	231.5	6.0	1116	1	SLPH_BACBR
9	213.5	5.5	1183	1	CNA_STRAU
10	211	5.4	1361	1	SPAP_STRMU
11	204	5.3	1053	1	SLPM_BACBR
12	201	5.2	1565	1	PAC_STRMU
13	199	5.1	941	1	GUN_BACSG
14	193.5	5.0	1091	1	NCAL_CHICK
15	187.5	4.8	1256	1	MRE_STRSU
16	187	4.8	944	1	IF2_HELPY
17	186.5	4.8	1004	1	SLPO_BACBR
18	185	4.8	1115	1	NCAL_MOUSE
19	184	4.7	644	1	NFM_RABIT
20	184	4.7	2464	1	MABP_MOUSE
21	182.5	4.7	728	1	TRDN_HUMAN
22	182	4.7	1220	1	IF2P_HUMAN
23	180	4.6	1500	1	SSPS_STRGN
24	180	4.6	1592	1	GTF2_STRDO
25	179.5	4.6	915	1	NFM_HUMAN
26	179.5	4.6	1827	1	NFM_MOUSE
27	178.5	4.6	848	1	NFM_MOUSE
28	178.5	4.6	998	1	SCA4_RICAK
29	178	4.6	1656	1	OMP_RICUA
30	177	4.6	1637	1	MSP_STRAU
31	176	4.5	845	1	NFM_RAT
32	176	4.5	1025	1	SLAP_CAUCR
33	175.5	4.5	933	1	SLAP_CAMPE

34	175.5	4.5	949	1	IF2_HELPY	Q9ZM46 helicobacte
35	175.5	4.5	4385	1	YP73_CAEEL	Q09222 caenorhabdi
36	175	4.5	1233	1	YF16_YEAST	P43597 saccharomyc
37	175	4.5	1654	1	OMP_RICRI	Q53047 r outer mem
38	174.5	4.5	1577	1	HLVA_PROMI	P16466 proteus mir
39	174	4.5	1433	1	REST_CHICK	Q42184 gallus gall
40	173.5	4.5	1790	1	USOI_YEAST	P25386 saccharomyc
41	172.5	4.4	1345	1	IGA3_HAEIN	P45385 haemophilus
42	172.5	4.4	1655	1	OMP_RICCN	Q9KKR3 r outer mem
43	172.5	4.4	1848	1	CBPA_CLOCL	P38058 clostridium
44	172	4.4	853	1	NCAL_BOVIN	P1836 bos taurus
45	172	4.4	1296	1	ASAL_ENTFA	P17953 enterococcu

ALIGNMENTS

RESULT	ID	SLAP_BACAN	STANDARD	PRT	814 AA.
1	AC	P49051			
DT	01-FEB-1996	(Rel. 33, Created)			
DT	01-FEB-1996	(Rel. 33, Last sequence update)			
DT	01-FEB-1996	(Rel. 33, Last annotation update)			
DE	S-LAYER PROTEIN PRECURSOR (SURFACE LAYER PROTEIN) (SURFACE ARRAY PROTEIN).				
GN	SAP.				
OS	Bacillus anthracis.				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;				
OC	Bacillus/Staphylococcus group; Bacillus.				
OX	NCBI_TaxID=1392;				
RN	[1]				
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 30-40; 209-218; 281-291 & 571-580.				
RC	STRAIN=STERNE;				
RX	MEDLINE=95138020; PubMed=7836294;				
RA	Etienne-Roumelin I., Stierd J., Duflot E., Mock M., Fouet A.;				
RT	"Characterization of the Bacillus anthracis S-layer: cloning and sequencing of the structural gene";				
RL	J. Bacteriol. 177:614-620(1995).				
CC	-1- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA.				
CC	-1- SUBCELLULAR LOCATION: CELL WALL.				
CC	-1- PTM: PROBABLY GLYCOSYLATED.				
CC	-1- SIMILARITY: CONTAINS 3 S-LAYER HOMOLOGY (SLH) DOMAINS.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib.ch).				
CC	-----				
DR	EMBL: Z6946; CA85408.1; -				
DR	InterPro: IPR00343; BIG_2.				
DR	InterPro: IPR001119; SLH.				
DR	Pfam: PF02368; Big_2; 1.				
DR	Pfam: PF00395; SLH; 3.				
KW	Signal; Glycoprotein; Repeat; Cell wall; S-layer.				
FT	SIGNAL	1	29		
FT	CHAIN	30	814		
FT	DOMAIN	33	93		
FT	DOMAIN	94	154		
FT	DOMAIN	155	213		
SQ	SEQUENCE	814 AA;	86620 MW;	C16380D26A1C6B101 CRC64;	

Query Match 99.0%; Score 3847; DB 1; Length 814;

Best local Similarity 99.0%; Pred. No. 1.5e-144;

Matches 777; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

1 AGKTFPPVADHMGIDISINVLKGAIVKGDGMFGRKELTFRAEATMAQTLNLPIDK 60

QY	DB	AGTFFPPDVPADHMGIDSTINILVYEGAVKAGNDKMGFEPEGKELTREAETATMAQLNLNPIDK	89
QY	30	DAKESFASOGOWMTPPIAAVEKAGVIGKTGNGEFPNGKIDRVSMASLVEAYKLDPTKN	120
Db	90	DAKPSFADSGOWMTPPIAAVEKAGVIGKTGNGEFPNGKIDRVSMASLVEAYKLDPTKN	149
QY	121	GTPATKFKRDELTLMWGKEKANILVELCISVGTGDMPEPKKTVYKAAAOFTIATDKOFGT	180
Db	150	GTPATKFKRDELTLMWGKEKANILVELCISVGTGDMPEPKKTVYKAAAOFTIATDKOFGT	209
QY	181	EAAVESAKATVTOQVEKFEKSAVEKLTKEKIDKTYNKANNDKVLVKFVTLSEBRSATVE	240
Db	210	EAAVESAKATVTOQVEKFEKSAVEKLTKEKIDKTYNKANNDKVLVKFVTLSEBRSATVE	269
QY	241	LYSNLAAKOYTVDVNKNVGTETAVAGSLEAKTIEMADQTVVADEPTALQFTVADENGTEV	300
Db	270	LYSNLAAKOYTVDVNKNVGTETAVAGSLEAKTIEMADQTVVADEPTALQFTVADENGTEV	329
QY	301	VSPGIEIEVTPPAEKINAKGKITLAKGTSTTVAAVYKKDGKVAESKEVAVSABGAAVAS	360
Db	330	VSPGIEIEVTPPAEKINAKGKITLAKGTSTTVAAVYKKDGKVAESKEVAVSABGAAVAS	389
QY	361	ISNMTPVAONKADFTSKDFKONNNVYEGDAAVYOVVELKDFNATTTKVEYESLNFPAV	420
Db	390	ISNMTPVAONKADFTSKDFKONNNVYEGDAAVYOVVELKDFNATTTKVEYESLNFPAV	449
QY	421	VDKATGKVTYLSAGKAPKVTYVSKSGKALVSHTEVEIEAFQAKAMKDIKLEKTIVALSTK	480
Db	450	VDKATGKVTYLSAGKAPKVTYVSKSGKALVSHTEVEIEAFQAKAMKDIKLEKTIVALSTK	509
QY	481	DVTDLKKAPAPLDQYKGFETAPVTVKVLBKDGKELKQKLEAKTVNRELYLMAAGDEAGN	540
Db	510	DVTDLKKAPAPLDQYKGFETAPVTVKVLBKDGKELKQKLEAKTVNRELYLMAAGDEAGN	569
QY	541	YTVVLTAKSGEKEKATLATELKPAGAFSKREYVGLTDELDRKYTEENOKNAMTVSVLPV	600
Db	570	YTVVLTAKSGEKEKATLATELKPAGAFSKREYVGLTDELDRKYTEENOKNAMTVSVLPV	629
QY	601	DANGVLTKGAEAEELKYTTNKEGKEVDATDAQVTVONNSVITVYGOGAKGETYKTVVL	660
Db	630	DANGVLTKGAEAEELKYTTNKEGKEVDATDAQVTVONNSVITVYGOGAKGETYKTVVL	689
QY	661	DGKLITTHSFVYVDTAPAKSLAVEFTSTLSKEVAPNADLKAALNLILSLSDGVPATTAKA	720
Db	690	DGKLITTHSFVYVDTAPAKSLAVEFTSTLSKEVAPNADLKAALNLILSLSDGVPATTAKA	749
QY	721	TASNVFVSADTNVAENGTVGAGATSIYKNTLVYDKCKEOKVEFDKAVOYAVASITKEA	780
Db	750	TASNVFVSADTNVAENGTVGAGATSIYKNTLVYDKCKEOKVEFDKAVOYAVASITKEA	809
QY	781	KPATK 785	
Db	810	KPATK 814	
RESULT 2			
SLAP_BACLI	STANDARD:	PRT:	874 AA.
AC	PA9052:		
DT	01-FEB-1996 (Rel. 33, Created)		
DT	01-FEB-1996 (Rel. 33, Last sequence update)		
DT	01-NOV-1997 (Rel. 35, Last annotation update)		
DE	S-LAYER PROTEIN PRECURSOR (SURFACE LAYER PROTEIN).		
OS	Bacillus licheniformis.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;		
OC	Bacillus/Staphylococcus group; Bacillus.		
OX	NCBI_TaxID=1402;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	SPRAIN-NM 105;		
FX	MEDLINE=97082965; Pubmed=8964497;		
TA	Zhu X., McVeigh R.R., Malathi P., Ghosh B.K.;		
RT	"The complete nucleotide sequence of the Bacillus licheniformis NM105"		

[illegible]

Db	103	YDALAAVYENGIFEGSVASTEEPAFKMQLTRSEAAKILYDAFLEBEGD----	ISEFADASTV	153
Qy	134	N-WGEKEMNIILVELGISVGTG-----	-----DOME	157
Db	160	KPKAKSYLEIAVANGVILKGSSEANGKTNLNPAPITTRQDFAVFVRTIEENVDAFPKVDKIE	1 : 1	219
Qy	158	-----PKRTVYKAEAAQFIATDKQFG-----	-----TEAAKVESAK	186
Db	220	VVDAKTLNVTLSDDGTKEVTLEKALIEPKRETEVEYFKIKDVEYKAKVTVYVTTATAVASVS	1 : 1	279
Qy	190	AVTQKVEYKRSKAVEKITKTREDIKVTNKKANDDKYLVEVTLSEBCKRSATVYLSNLAQK	24	459
Db	280	ATNLKEVVAEEDGTVDKETAED--AANYALKSGKITSVSLAANAKTAATVILDTKLNNK	33	737
Qy	250	TYTVDVN--KKGKTEVAAGSL-----EAKTIEADQTVAADEH-----TALQF	290	390
Db	338	ADALISINWKAQDEINKANVEFPAVDNKKIPEYIEVYKSLGKAKAVYLSEHVEVLSSTNF	39	737
Qy	291	TVKDE-----NGTEVVSPEGIEEYTPAAEKINAKGEITLNGSTSTTVKAVYKKD	335	435
Db	398	TLDGAYVAGNYVMGAGNTVILTFYSSSALSVGHKILTVSGAKPAGFVSLNSTHEPK--	455	511
Qy	340	KGYVAESKEVYVNSAGVAASISWYVAEQKAD-----FTSKDPKQNNKYYE--GD	38	737
Db	456	---VEDEKADPTVEATPTLETTLTFTSESDIDMPTVKASANYWYKSGSKKASEFERIAD	51	737
Qy	390	NAYVOVELKDDQFNNAVTTGKVEY-----ESLNTFEAAVDKATGKVTYLSAQAQKVTYVK	44	737
Db	513	NKYKVV-FKSGEKLTPGKQVDDYVEDIKDYSDNKIADKTATVPELIDQTPREVRKKTAL	57	737
Qy	444	DSKG-KALVSHTVIEIAPFAQKAMKDILEKTVALSTKDYTDLK-VKAPVLDQYGEFTA	50	737
Db	572	DEKTIKVFSTKTVDGESAIFKTGNVTYKDKDKVYSDVTVDSKDSKSVIIDLTSKYSVG	63	737
Qy	502	PVTVKAVLD-KQCKELKEKLEAKYVNRVLVYNAAGQEGNGTIVVLTAKSGKEKAKATLAL	56	737
Db	632	ENTTIKIKVKKATKILNNTMLD--TYGK--FTTRDKCEPDYEHYIMADA--KAKVVLKF	68	737
Qy	561	ELKAPGA---FSFEYVRLGDT--ELDKYVTEENQKNAMTVSLP-----VDANG	60	737
Db	665	DKKMAASLADYSNVLKVINDTLQTLSDVATLSVSNATVVTITFPAETIGDDVFPASG	74	737
Qy	605	LVLKGA---EAAELKV---TTTNKEGEVDATDAQVTVONNSVITVGGAKKAGETY--K	65	737
Db	745	KAISGSGKVYNNELQVMGVKDTSGVHKHFKNGSEKKITLSTST--TPLKLAKIDKDYDAK	80	737
Qy	656	VTV-VLDDKLLTTHSEKVVDDPAPRAKGLAVEFTSTSLK-----EYAP	69	737
Db	803	YTAELVDKRTKYVKEFSTYINBA---AANFTSESHIDSIOVNGSTVYVYKFKDEINT	85	737
Qy	697	NA---DLKAALINLISVDGVNAT--TAKATASNEFEVADINVAENGVTAQKAGTSIYV	75	737
Db	858	NASDDLKVLNLSKLVDINGNESTNNTPTAIRAGINLDSVAPVYVGEPVYKKEITLTFSS	91	737
Qy	752	KNLTVVKKDGKEQKVEE 767		
Db	918	ENLTSVSIGEVLSTDF 933		
RESULT 4				
SLAP_ACEKI STANDARD; PRT; 762 AA.				
AC P22258:				
PT 01-AUG-1991 (Rel. 19, Created)				
PT 20-AUG-1991 (Rel. 19, Last sequence update)				
DE CELL SURFACE PROTEIN PRECURSOR (S-LAYER PROTEIN).				
OS Acetogenium kivui.				
OC Bacteria; Firmicutes; Bacillus/Clostridium group;				
CC Thermotoga bacterioides group; Thermotoga bacterioides.				
OK NCBI_TaxID=2325;				
RP [1]				
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				

RC STRAIN-DSM 2030; Pubmed=2681162;
 RA MEDLINE=90036724; PubMed=2681162;
 RX Peters J., Peters M., Lottspeich F., Baumeister W.;
 RT "5-layer protein gene of Acetogenium kivui: cloning and expression in
 Escherichia coli and determination of the nucleotide sequence.";
 RL J. Bacteriol. 171:6307-6315(1989).
 RN [2]
 RP PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITES.
 RC STRAIN-DSM 2030;
 RX MEDLINE=92281680; Pubmed=1596358;
 RA Peters J., Rudolf S., Oschkinat H., Mengele R., Sumper M.,
 RT Kellermann J., Lottspeich F., Baumeister W.;
 RT Evidence for tyrosine-linked glycosaminoglycan in a bacterial
 surface protein.";
 RL Biol. Chem. Hoppe-Seyler 373:171-176(1992).
 RN [3]
 RP DOMAINS.
 RX MEDLINE=94156823; Pubmed=8113161;
 RA Lupas A., Engelhardt H., Peters J., Santarius U., Volker S.,
 RA Baumeister W.;
 RT "Domain structure of the Acetogenium kivui surface layer revealed by
 electron crystallography and sequence analysis.";
 RL J. Bacteriol. 176:1224-1233(1994).
 CC -1- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY
 OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA.
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
 S-LAYER WITH HEXAGONAL SYMMETRY.
 CC -1- PTM: THE CARBOHYDRATE CONTENT OF THIS PROTEIN IS ABOUT 8% WHICH
 CORRESPOND TO ABOUT 40 TO 50 SUGAR MOLECULES PER MONOMER. O-LINKED
 GLYCANS CONSIST OF GLC, GALNAC AND GLCNAC.
 CC -1- SIMILARITY: CONTAINS 3 S-LAYER HOMOLOG (SLH) DOMAINS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M31069; AAA21930.1; -.
 DR PIR: A34355; A34355.
 DR InterPro: IPR001119; SLH.
 DR Pfam: PF00395; SLH; 2.
 DR PROSITE: PS01072; SLH_DOMAIN; 2.
 KW Signal; Glycoprotein; Repeat; Cell wall; S-layer.
 FT SIGNAL 1 26
 FT CHAIN 1 762
 FT DOMAIN 27 762 CELL SURFACE PROTEIN.
 FT SLH 1.
 FT SLH 2.
 FT SLH 3.
 FT DOMAIN 146 204 SER/THR-RICH.
 FT DOMAIN 473 479 SER/THR-RICH.
 FT DOMAIN 625 630 SER/THR-RICH.
 FT CARBOHYD 297 297 O-LINKED (GLC. . .).
 FT CARBOHYD 516 516 O-LINKED (GLC. . .).
 FT CARBOHYD 520 520 O-LINKED (GLC. . .).
 FT CARBOHYD 632 632 O-LINKED (GLC. . .).
 SO SEQUENCE 762 AA; 83785 MW; 34EC9C784DECA67E CRC64;
 Query Match 6.68; Score 255.5; DB 1; Length 762;
 Best Local Similarity 22.7%; Pred. No. 0.0014;
 Matches 177; Conservative 89; Mismatches 275; Indels 239; Gaps 35;
 Oy 5 EPOVPADHWGIDSYNYLVGAVGNGDKMPEPKELTRAAAMMQLVLPIDKDAKP 64
 Db 96 FKDPVPHMAVGQINLAKGLAOGVNGKFDPSLRYAQLAFVLRALGF---KDL-- 150
 Oy 65 SFADSOQWYTPFLAAVEKGVINGTNGEPNGKIDRVSNASLLVEAYKLDTRVNGTPA 124
 Db 151 -----DWPRGYLAKKODGLVGLNLA--NGVIRKGDALLLDRLALEV-----PM 194
 Oy 125 TKRQDLETLMMGKEMKANIIVELGISVGTGDOMEPKKTVTKAAAOFLAKTDKQGTGEAK 184

Db 195 VKYVD-----GKE-----VLGEPLIS-----KVATKAEYT-VIAFNAQDRSVEGK 234
 Oy 185 V-----ESAKAVTQKVEKFSKAVEKLEKEDIKVTNNKNNKVLVKE-----VTLSEKRR 235
 Db 235 VAUDDKDKGLTTINAGLVDFSEYLGK---KYIYSEKFGCPYVYAEQNDVNSTEBCQD 290
 Oy 236 SATVELYSNLAQOTYVDVN-----KVQG-----TEVAV-----GSLFA 270
 Db 291 SVGTYYKNDNDKTAIKVDNNAVLYNGYLKFSKVTEKAGEAVTIINNLYIANGSYDN 350
 Oy 271 KTIEMADQTVVADEPTALQFTVKDENGTEVVSPECIEVTPAAEKINKGETTAKGS- 329
 Db 351 STI-----VYNDVQSGKRYLNRSN-----YELKGYTVYGAVSK 385
 Oy 330 -TTVKA-----VYKDGKRV-----AESKEVKYSAAGAVALSISNTV 366
 Db 386 VTDIKANDYIYKQDYNGVNGVYIVYRNQVGTGYTEKSVSGSTYKASIDNVS-YTV 444
 Oy 367 AEONKAD-----FTSKD-----FKONKYEEDNAVYQVEL 397
 Db 445 ADNVMWNLDEPKKVTYLNKNDNVIGISSTTTTAVVYALFKESDPTAMFAKVLIL 504
 Oy 398 KDQ-----FNAVTTGKVEYESLN-----TEVAVYDKATGVTVL-----SAGKAPK 439
 Db 505 PDAAEKVFDAYSD--VYDKVNLAEGITVYTVYDANGKLNIDQRANDPFSSASAKADAK 562
 Oy 440 VTVAQSKKALVSHVTEIE-AFAQAKMDIKLEKTNAVALSTKDPDLDKAPVLDQY-- 495
 Db 563 VLTEGSTYYITDNTVILNNTSDGKALKITDL-----KATINLVNI-VADNNV 612
 Oy 496 -----GKEFTAPYTVKVLDDKGELEKQLEAKYVRELVLNAAQAGNVTYVITAK 548
 Db 613 AKVYFVNASVSTTSHV-----AYVFGADVYNGSTFRLFLVLENGQ 658
 Oy 549 SGEENAKRTLLLEKBAFASKFEVGRGDTELDRTYTEENKNNMTVVLVYDANGVLK 608
 Db 659 TKTYDANAQLATNV-----THKAVVLTJTNKIANI--ALPTVASGVKLT 701
 Oy 609 GAFAELKVTYTNKEKFEVD-----ATDAQVTYQNNSVITVVGQAK-----AGEYKRVV 659
 Db 702 NIDQANLRITDTTNKGYILDPNFIVVDNNGNLKGLSDTKTGQVLYTNVQKGVVITIV 761
 RESULT 5
 SLAP_BACST STANDARD; PRT; 1228 AA.
 AC P35825;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE S-LAYER PROTEIN PRECURSOR (SURFACE LAYER PROTEIN).
 GN SBA.
 OS Bacillus stearothermophilus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Geobacillus.
 OX NCBI_TaxID=1422;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PV72;
 RX MEDLINE=94320770; Pubmed=8045409;
 RA Kuen B., Sleytr U.B., Lubitz W.;
 RT "Sequence analysis of the sbas gene encoding the 130-kDa
 surface-layer protein of Bacillus stearothermophilus strain PV72.";
 RL Gene 145:115-120(1994).
 CC -1- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY
 OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA.
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
 S-LAYER WITH HEXAGONAL SYMMETRY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its

Oy		628	DATDAQVTQNNNSYITWGGAKAGEFYKVVTVLDG-----KLITTHSEFVVDTPAP	6779
Db		957	-TKKFPTPSQELKAGTIVSYTL--DGVDRKGNTISKIIT--SKFTVSANP	1000

FT	SIGNAL	1	35	POTENTIAL.
FT	CHAIN	36	1861	AMYLOPULULANASE.
FT	DOMAIN	928	1018	FIBRONECTIN TYPE-III 1.
FT	DOMAIN	1157	1248	FIBRONECTIN TYPE-III 2.
FT	ACT_SITE	628	628	BY SIMILARITY.

Query Match 6.0%; Score 231.5; DB 1; Length 1116;
 Best Local Similarity 21.6%; Pred. No. 0.02;
 Matches 193; Conservative 125; Mismatches 307; Indels 269; Gaps 47;

```

OY 17 STNYLVKGVANGNKGMEPEKELTRAEATMMQILMLPLDKDK-----PSFDSOG 71
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 70 TYKRLDALGLVAGYNGDGDADKTTIRAEFATLIVARGL--EÖGAKLAOFNTTYDVNS 127
OY 72 -OMYTFPIAAVKAQVING-TGNGEPENKIDRVSMASLVEAYKADTVNGTPATKFXD 129
      ||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 128 TDMFAGFVAVASGEELVYKFPDKSFRQNOVYAEAVTMIVRALGEPYRGV----- 180
OY 130 LETLMMGKKANILVELGISVGTGDOMEPRKVTYKAAOFLAKTDKÖGTEAAKVESAK 189
      ||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 181 -----WPSNMISKGESELTAKG-----INNPMOOFATIRK----- 212
OY 190 AVTTQKVEKFSKAVEKLEKEDIKVTNK---ANNDKVLVEKVSLSBK--RSATVELYSN 244
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 213 -MIDNALRKMLKEQIEFYGDIRLNTDETLTKYLVYRDMQMAHEKGNSEDELFLVTN 271
OY 245 IAAKÖTYTVDNKVGKTEVAVGSLKAKTTEMADQTV--VAD--EPTA-----LOFTVKDE 295
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 272 VPAIGLSLKAMBV-----TLNGKADADLSNTTYKVAEGINPAPGQKQVWIMKD 323
OY 296 NGTEVAVSPGIEFVTPAAKINA--KGETTLAKGSTTVKAVYKKDKGVAVESKEVKS 352
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 324 RENVIYMMGSEDEEDVYMDRVSAIYLGKRAF-----TDDIVKDLKSD-----LDDVKIE 373
OY 353 AEGAAVASISNMVVAEQNK-----ADPESKDFPKN-----KYEGDN---AYVQ 394
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 374 MDG-----SEKSTRJEDTYITTFNFTFNDPVDLSKITIDNDFGVKVLNDNEVAYLH 429
OY 395 VELKQFNAVTGKVEYESLNTAEVAVDKATGKVTYLSAGK----- 435
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 430 I-IDQOTDKSVKGVYGS--KVISKIDADKKITNLDNSKSFEDLDQDCKFLVELDG 486
OY 436 APVKV-TVVDS-----KGRALVSHYVEIEAPQAKMKDILEKTNALSKRDVT 483
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 487 QPAKLDLDESVDYVYVADGDKXYLV-----FANENVAEGVYK-----VVSINKT 534
OY 484 DLKVAKPVLDQYQKEPTAPVTVKVLDDKDEKLEKQLEAKVYR--ELVYNAGQEGN 541
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 535 DIR-----LTVGK-----TYKYV-PDASYSSENANKDVAKVSDLDLSLNGEYV--- 579
OY 542 TVVLTAKSSEKAKATLALAKAPAF-----SKFEVRGLDTE 579
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 580 KLLDPSGVRHIETKDAIDRRPLAITRGATYNSKDTYFTVMÖKKKTÖIVSLDK 639
OY 580 --LDKYVTEENOKNMTVSVLPVDANGLVKGAELKVTTNKEGKEYDATDAQVTVQ 637
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 640 DIYDRGVNYDKSNDK-----RQAFKDELVELLOPKVVKEDSATDAMQTV- 684
OY 638 NNSVITVGGAGAGETKYTVVLDGKL-----IT--TISFKV 672
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 685 ---LLEVNFDK-GEVDKYK-VLDSKLKSEKSTWKDLADEDDVGVDEVTOKTAVFKM 739
OY 673 V-DTAP---TAKGLAEVFSTSLKEVAPNADL-----KAALNITISVDCVPATTAK 719
      ||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 740 TGDILPATGKTKGELKNAGTAKKDVAKKSDKLVMVSVDEDDKEVQALIFVDDG-----S 793
OY 720 ATASNEVEVSADTNNVAENGTVGAKGATSIYKNLTVVDG-----KEQKVEFD 768
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 794 GLGDHGF-----GMVKÖYGTASKÖDTIT-----VTKGDGVSVERKEYKLDDG 836
  
```

DE COLLAGEN ADHESIN PRECURSOR.
 GN CNA.
 OS Staphylococcus aureus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Staphylococcus.
 OX NCBI_TaxID=1280;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FDA 574;
 RX MEDLINE=92165839; PubMed=1311320;
 RA Patil J.M., Jonsson H., Guss B., Switalski L.M., Wlberg K.,
 RA Lindberg M., Hoecek M.;
 RT "Molecular characterization and expression of a gene encoding a
 RT Staphylococcus aureus collagen adhesin."
 RL J. Biol. Chem. 267:476-477(1992).
 RN [2]
 RP ERRATUM.
 RA Patil J.M., Jonsson H., Guss B., Switalski L.M., Wlberg K.,
 RA Lindberg M., Hoecek M.;
 RL J. Biol. Chem. 269:11672-11672(1994).
 RN [3]
 RP COLLAGEN-BINDING DOMAIN.
 RC STRAIN=FDA 574;
 RX MEDLINE=94032261; PubMed=8218209;
 RA Patil J.M., Boles J.O., Hoecek M.;
 RT "Identification and biochemical characterization of the ligand
 RT binding domain of the collagen adhesin from Staphylococcus aureus."
 RL Biochemistry 32:11428-11435(1993).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ÅNGSTRÖMS) OF 169-318.
 RX MEDLINE=97475225; PubMed=9334749;
 RA Symersky J., Patil J.M., Carson M., House-Pompeo K., Teale M.,
 RA Moore D., Jin L., Schneider A., Delucas L.J., Hoecek M.,
 RA Narayana S.V.L.;
 RT "Structure of the collagen-binding domain from a Staphylococcus
 RT aureus adhesin."
 RL Nat. Struct. Biol. 4:833-838(1997).
 CC -I- FUNCTION: MEDIATES ATTACHMENT OF STAPHYLOCOCCAL CELLS TO
 CC COLLAGEN-CONTAINING SUBSTRATA.
 CC -I- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. CELL WALL.
 CC -I- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
 CC IN THE REGION OF THE MEMBRANE ANCHOR.
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M81736; AAA20874.1; -
 DR PDB; 1AMX; 24-JUN-98.
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR PROSITE; PS00343; GRAM_POS_ANCHORING; FALSE_NEG.
 KW Signal; Repeat; Transmembrane; Cell wall; 3D-structure.
 KM
 FT SIGNAL 1
 FT CHAIN 30 1183
 FT DOMAIN 30 1157
 FT TRANSMEM 1158 1177
 FT DOMAIN 1178 1183
 FT DOMAIN 151 318
 FT DOMAIN 533 1093
 FT DOMAIN 1093 1157
 FT DOMAIN 1151 1156
 FT REPEAT 533 719
 FT REPEAT 720 906
 FT REPEAT 907 1093
 SQ SEQUENCE 1183 AA; 133066 MW; B6A1C0072E575D76 CRC64;

Query Match 5.5%; Score 213.5; DB 1; Length 1183;

Best Local Similarity 22.0%; Pred. No. 0.11;
Matches 199; Conservative 109; Mismatches 304; Indels 291; Gaps 47;

```

OY 19 NYLVEKGVKNDKGMFPGKELTRAEATMAQILNP-----IDKDA 62
DB 232 NYVSGQSAITDFEKNF--PGSKITVDNTRKNTID--VTIPQGVGSYNSFKTKITNQ 287
OY 63 KPSFADSGOWYTPP-----IAAVERKAGVYKGTNGEPNGKIDRVSMASLL 109
DB 288 OKEFVNNSQAOYQEHGKEVEVNGKSPNHTVHNINANAGLEGTVKGELKVLKDDKDKAPIA 347
OY 110 VEAYKLDTRVNGTPATKREDLETLMWGKEMANILVELISVGTGD---QMEPKK--TVT 163
DB 348 NVKFKLSKKDGSVVKDQNKQEIITDANGINIK-----ALPSGDYILKEIARPPYFD 402
OY 164 KAAEAOEIAK-TDKQ--FGT--EAAKVESAAVTTQKV-----EYK----- 199
DB 403 KDEYPTFMKDTQNOGQFTTIEENAKAIEKTKDVSQOKWEGSTQKPTIYFKLYKQDDNQ 462
OY 200 ---FSKA-VEKLTKEKEDKVT--NKAND-----KYLVREVTLSDEKRSATVELY--- 242
DB 463 NTTVDKAEIKKLEDGTTKTYMSLDPENDKKAIKYLVKVNMOGE--DTPPGYTRKE 520
OY 243 -----SNEAAQTYTVDNKYCK--TEVAVGSL---EAKTIEMADQ--- 279
DB 521 NGLVVTNTERPIETTSISGEKVMDDKQNOGDKRPEKVSYNLLANGKKTLDVTSETNMK 580
OY 280 -----VVADEPTLOFTVDE-----NGTEVVSPEGIEVVTPAEKINAKGETTL 324
DB 581 YEFKDLPRYDEGKKIETVYEDHVKDYTTDINGTITNKK-----YTP-----GE--- 624
OY 325 AKGTSTTVKAVNY-----KKDKVNAESK-----EYKVSAGAANAISNMT-----VAEON 370
DB 625 ---TSATVTKNMDDNNQDGRPEIEIVELYQDGAATGATATILNESNNWHTWTGLDLEKA 681
OY 371 KADFTSKDFQNNKVKYEGDNAY--VOVELKQDFNAVTGKV--EYESINTEVAVVDKATGK 427
DB 682 KGOOVKTYVEELTRV---KGYTHVDNNMGNLITNKKYTPETTSISGEKVMDDKDN-- 735
OY 428 VTVLSGKAPVYTVK--DSGKALVSHTVLEI EAFQAKAMDI-----KLEKTNVALST 479
DB 736 ---QDGRPEKVSYNLLADGKVKKTLDTVSETNMKYEFDLPKYDEGKKIETVTEHDV 791
OY 480 KDVTDLKVKAPVLDQYGEKFTAPVTVKVLDP---KDGKELKEOKLE---AKVYNRELVL 531
DB 792 KDVTDTINGTITNKKYTPETTSATVTKNMDDNNQDGRPEIEIVELYQDGAATGATATIL 851
OY 532 NAAQGEAGNTVVLTAKSGEKKAATLALAKAPGAFSKP-----EVRGLDTELD--- 581
DB 852 N---ESNNWHTWT-----GLDEKAKGOQVKYTVLEELTRKVGYYTHVDNNM 895
OY 582 -----KYVTE-----ENO--KNAMTVSVLPVDANGVLVKCAEAELKYTT 619
DB 896 GNLITVNNKTYPETTSISGEKVMDDKQNOGDKRPEKVS---NLLANEKVKTTLDVTS 949
OY 620 -TN-----KEGKEVDAT-----DAQVTVNNNSVI 642
DB 950 EFNMKYEFKDLPRYDEGKKIETVTEHDVHKDYTTDINGTITNKKYTPETTSATVTKNMD 1009
OY 643 TVGQGAKEGTYKVVVLDGKLI-----TTHSEKVVDTAPATKALAEFTSTSL 691
DB 1010 NNNQDGRPEIEIVELYQDGAATGATATILNESNNWHTWTGLD--EKAKGOQVKYTVDEL 1067
OY 692 KEY 694
DB 1068 TRV 1070

```

```

DT 01-NOV-1991 (Rel. 20, Last sequence update)
DE 01-FEB-1996 (Rel. 33, Last annotation update)
DB CELL SURFACE ANTIGEN I/II PRECURSOR.
OS SPAP.
OC Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [11]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=NG5 SEROTYPE C;
RX MEDLINE=90076473; PubMed=2687020;
RA Kelly C., Evans P., Bergmeier L., Lee S.F., Proguliske-Fox A.,
RA Harris A.C., Aitken A., Bleiweis A.S., Lehner T.;
RT "Sequence analysis of the cloned streptococcal cell surface antigen
RT I/II".
RL FEBS Lett. 258:127-132(1989).
RN [12]
RP SEQUENCE FROM N.A.
RC STRAIN=NG5 SEROTYPE C;
RX MEDLINE=91207143; PubMed=1982405;
RA Kelly C., Evans P., Ma J.K.C., Bergmeier L.A., Taylor W., Brady L.J.,
RA Lee S.F., Bleiweis A.S., Lehner T.;
RT "Sequencing and characterization of the 185 kDa cell surface antigen
RT of Streptococcus mutans".
RL Arch. Oral Biol. 35:335-385(1990).
CC -1- FUNCTION: SURFACE PROTEIN ANTIGEN IMPLICATED IN DENTAL CARIES.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -1- SIMILARITY: BELONGS TO THE SPAP/SPB/SPAA FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X17390; CAA35253.1; -.
DR PIR: S06839; S06839.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00746; Gram_pos_anchor.
DR PROSITE: PS00343; GRAM_POS_ANCHORING.
KW Antigen; Signal; Transmembrane; Repeat; Cell wall; Dental caries.
FT SIGNAL 1 38
FT CHAIN 39 7
FT CHAIN 997 1561
FT DOMAIN 39 1536
FT TRANSMEM 1537 1561
FT DOMAIN 1557 1561
FT DOMAIN 1557 1561
FT DOMAIN 60 550
FT DOMAIN 219 464
FT DOMAIN 847 963
FT DOMAIN 1528 1533
SQ SEQUENCE 1561 AA; 170060 MW; 540D92768FC8AB4B CRC64;

```

Query Match 5.4%; Score 211; DB 1; Length 1561;
Best Local Similarity 22.2%; Pred. No. 0.19;
Matches 181; Conservative 100; Mismatches 302; Indels 232; Gaps 39;

```

OY 25 GAVKND-----KGFEPGKELTRAEATMAQILNP-IDKDAKPSADSGOWYTP 76
DB 53 GTGTGNPATNLPRAQGSQAQADQSQTLEKRWVHTVEVPTDIDAADAKAKSAG----- 107
OY 77 FIAAVERKAGVYKGTNGEPNGKIDRVSMASLLVEAYKLDTRVNGTPATKFKDLETLMWG 136
DB 108 -VAVVQDADADVNNKGTVTAE-----EAVQKETELEKEDYTKQAEIDIKK----- 147
OY 137 KEKANILVELGISVGTGQDMEPKRYTKAEAAOIFAK---TDKQFG-----TEAKVE 186

```

```
Db 148 -----TTDOYKSDVAHAEVAFIKAKNQATKQYGDWVAHKAUEVERIN 192
OY 187 SAKAVTTOREVEVFSK-----AVEKLTKEKIDKVTNKANDKVLVEVTLSEDKRSATVE 240
Db 193 AANAASKTAYEAKLAQYADLAIVOK-----TNAANDASQOKLA--AYQAELEKRVGE-- 243
OY 241 LYSNLAOKOTY--TVUNVKYKTEVAVGSLAEKTIEMADQTVVADEPTAL-QETVVDENG 297
Db 244 --ANAAKAAAYDTAVANNAKNTAIAAANE---IRKRNATAKAEYETKLAQYQAELEKRV 298
OY 298 TEVVSSEGIEPTVPAEKINA--KGEITLKGSTTYKAVYKKGKGVAAVESKEVKSVEGA 356
Db 299 QEANAANEADY---OAKLTAIYOTELAROVKANAADAAL-----EAAVAANNA 343
OY 357 AVASISNMTVAONKADFTSKDFKONNKVYEGDNAYOVLEKDFNAVTTGKEYE----- 412
Db 344 KNAALTAENTATIKORNEKATKATEALQYEAADLAIVK-----KANAANEADYQAKLT 396
OY 413 SLNTEVAVVADKATGVTVLSAGKAPVK--VTVKDSKGLALVSHTEIEA--FAOKAMKDI 468
Db 397 AVOTELAROVKAN-----ADAKAAVEAAVAANNAALTAENTATIKRNADAKADYEA 450
OY 469 KLEK--TNAVALSTKVDTLKVKAPVLDQYKFTAPYTVAVLDKDEKLEKLEAKYVN 526
Db 451 KLAKYQADLAKYQKDLADY-----PVKLAYEDEQASIKAAALAE----- 489
OY 527 RELVNAAGOEAGNVTYVTLFAKSGEK-----EAKATLAE-----LKAPG--AFSKFEV 573
Db 490 -----LEKHNEDGN-----LTPEASNOLVYDLEPANSLTTDGEFLKASAVDAFSEK--- 538
OY 574 RGLDELKDYTEENOKNMAVTSLPYDANGVLKGAEEALKYTTNKGKEVEDATDAQ 633
Db 539 -----STSKAKY-----DOK-----ILOLD--DLDTTNLEQSNDAVSMELXGNGDKAGWS 583
OY 634 VTYQONS-----VITVGGAAGETRYKYTVVLDGKLITTHSK--VYDPAFTPKGLAV- 684
Db 584 TYVSNNSOVKMGSVLLERQGSARATYTNLONSYYNGKIRIKYKYVYDPRSKFGQKQW 643
OY 685 --EFTSTSLKEVAPNADLKAALINTLSVDGPATTAKTASNEVFASVDNVAENGTVG 742
Db 644 LGFTDPTL-----GV-----FASATYQGVKKN----- 666
OY 743 AKGATSIYVKNLTVKDGKEQKVEPKAVOVAVSI 777
Db 667 -----TSIFIKNEFTFYDEDEGKPIFNDAALSVASL 697

RESULT 11
SLPM_BACBR STANDARD: PRT; 1053 AA.
AC P06546;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE MIDDLE CELL WALL PROTEIN PRECURSOR (MWP).
OS Bacillus brevis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Brevibacillus.
OX NCBI_TaxID=1393;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=47;
RA MEDLINE=8115203; PubMed=2828336;
RA Tsuboi A., Uchihl R., Adachi T., Sasaki T., Hayakawa S., Yamagata H.,
RA Tsukagoshi N., Uda S.;
RT "Characterization of the genes for the hexagonally arranged surface
RT layer proteins in protein-producing Bacillus brevis 47: complete
RT nucleotide sequence of the middle wall protein gene.";
RL J. Bacteriol. 170:935-945(1988).
RN [2]
RP SEQUENCE OF 1-199 FROM N.A.
RC STRAIN=47;
```

```
RX MEDLINE=87137282; PubMed=3029027;
RA Yamagata H., Adachi T., Tsuboi A., Takao M., Sasaki T.,
RA Tsukagoshi N., Uda S.;
RT "Cloning and characterization of the 5' region of the cell wall
RT protein gene operon in Bacillus brevis 47.";
RL J. Bacteriol. 169:1239-1245(1987).
RN [3]
RP SEQUENCE OF 676-1053 FROM N.A.
RC STRAIN=47;
RX MEDLINE=87008404; PubMed=2428810;
RA Tsuboi A., Uchihl R., Tabata R., Takahashi Y., Hashiba H., Sasaki T.,
RA Yamagata H., Tsukagoshi N., Uda S.;
RT "Characterization of the genes coding for two major cell wall
RT proteins from protein-producing Bacillus brevis 47: complete
RT nucleotide sequence of the outer wall protein gene.";
RL J. Bacteriol. 168:365-373(1986).
RN [4]
RP SEQUENCE OF 1-50 FROM N.A.
RC STRAIN=47;
RX MEDLINE=90078123; PubMed=2512285;
RA Tsuboi A., Uchihl R., Engelhardt H., Hattori H., Shimizu S.,
RA Tsukagoshi N., Uda S.;
RT "In vitro reconstitution of a hexagonal array with a surface layer
RT protein synthesized by Bacillus subtilis harboring the surface layer
RT protein gene from Bacillus brevis 47.";
RL J. Bacteriol. 171:6747-6752(1989).
CC -1- FUNCTION: THE MIDDLE WALL PROTEIN BINDS TO PEPTIDGLYCAN AND TO
CC THE OUTER CELL WALL PROTEIN.
CC -1- SUBUNIT: THE MIDDLE CELL WALL LAYER IS COMPOSED OF SUBUNITS OF
CC THE MIDDLE CELL WALL PROTEIN. THESE PROTEINS FORM A HEXAGONAL
CC ARRAY WITH A LATTICE CONSTANT OF 14.5 NM IN THE MIDDLE CELL
CC WALL LAYERS.
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC HEXAGONAL S-LAYER.
CC -1- SIMILARITY: CONTAINS 3 S-LAYER HOMOLOGY (SLH) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M15364; AAA87321.1; -
DR EMBL: M14238; AAA22372.1; -
DR EMBL: M19115; AAA22760.1; -
DR EMBL: M31828; AAA22619.1; -
DR PIR: A25039; A25039.
DR PIR: A28555; A28555.
DR InterPro: IPR001119; SLH.
DR Pfam: PF00395; SLH; 2.
DR PROSITE: PS01072; SLH DOMAIN; 2.
KW Cell wall; S-layer; Signal; Repeat.
FT SIGNAL 1 23
FT CHAIN 24 1053 MIDDLE CELL WALL PROTEIN.
FT DOMAIN 29 92 SLH 1.
FT DOMAIN 93 143 SLH 2.
FT DOMAIN 144 203 SLH 3.
SQ SEQUENCE 1053 AA; 117146 MW; DB421318BD9D5E4F CRC64;

Query Match 5.38; Score 204; DB 1; Length 1053;
Best Local Similarity 21.0%; Pred. No. 0.23;
Matches 187; Conservative 128; Mismatches 348; Indels 228; Gaps 43;

OY 4 TFPVPPADHMDISINILVKGAVKGNKMGFEQKELTRREAATMMQIINTLPIDKAK 63
Db 93 TYTVQKSTWDPAQFVNVAASGEIVKGPDKSKFQNOVTVAAVMTVRLAGY-----E 146
OY 64 PSFADSGOWTTPFIAVEKAGVIKGTGNGFEPNGKIDRVSMASILVAVYKIDTK----- 118
Db 147 PSV---KGVWPSNMSIKRASELNTARSITT---PNNATRGDIFKKLDNALVADLMEGYEF 200
```

```

OY 119 -----VNGPATKRF-----KOLETLNMGKEKANIIVEIGI-----SVGTGDOMERK 160
DB 201 GTDIRHEITKETLLTKLTKVTWRME---MAOEGNSEDDLPJLTNPAGLGK-----251
OY 161 TVTAEAAOFIAKTDOGFTEAAKVESAKAVTTOKEVEKFSKAVEKLEKEDIKVTNRANN 220
DB 252 --IKANEVTLNGK-DAGIGNTTYK--ADGINANDPQOHQVMIKDKDEVIYMESEST 306
OY 221 DKVL-----VKEVTS-----EDKRSATYELVSNL-----AAKOTYTVDNKVKTEVA---264
DB 307 DOEIVMDRVGEFTLKGTFEDPKDLSNSDLADLELDAESEKSYRFNNK---TKVTYNF 362
OY 265 -----VGSLEAKTIEMAD-----QTVVADEPTALOP--TVKENGSEVVSPEIEVVT 310
DB 363 TRENDPVDGLKEITIKDNADSGFTGCAKVVLDNNNEIAYIHVIDQSNMK--EDEGVKYS 420
OY 311 PAAEKINA-KGEIT-----LAAGTSTYKAVYK 337
DB 421 EVISKIDTDKKKITNRNDKFNLDGKEGKDFLVFLNGKPAKFSDLKEG---MYVSYY 477
OY 338 KDKVVAESKEVKYASAGAIVASISNMTVAQNKADFTSKRFKONKYYEG-----DNA 391
DB 478 ADGD---EDKLIVFATDTVVEGKVDK--VVSRRNNDRFLTIGDTRYRVEGATESDGNK 532
OY 392 YVOVELKDOENAVTTGKVEYESLNTG--VAAYDKATGKVTYLSAGAKPVYKDSKSKAL 450
DB 533 DVODIDKDHMDLV-----DSLDETVKLYLDASGRVHIE-----TKDAIDRRKQKAI 580
OY 451 VSHIVLELEAQAOKMAKDIKLEKTVALSTKDVYDLKAKAPVLDQYGEFTAPVTVK-VLD 509
DB 581 VTRSAFVN-----TSKDTWDFR---VLTQKGEITVSLKAKNIYD 617
OY 510 KDKG-----ELKEOKLEAKYVNR---ELVNLAAQOAGNYTVLTKSSEKEKA 556
DB 618 FDKGNFSROKNODDLEDIIVPSKDKDTLLEVTYLDADG-KPGKVEELKPVYKQJESKA 676
OY 557 TLALLEKAPCAFSEFEVR-----GLDTELDKVYTEENOKNAMTVSVLPVDANGSLVKA 610
DB 677 WDLADEDDDMVDGYEVTDKTAVFNMTGLEESSEKRELKNAKPAKFDV-----A 727
OY 611 EAELKVTNTNKEGKEVDATDAQVTVQNNSVIT-----VGO-----GAAGETVYKTVVLDG 662
DB 728 DENDLSIYTVNOKDEVEA---IFVEGDLGTDAGHYGVIDGRRGGKDTIRVWERKG 783
OY 663 KLITTHSFVY-----VDTAPFAKGLAVEFTSTSLKEV-----APNADLKALLNLS 709
DB 784 DKYVEKEVKLDGDDDLKDEDIRNDIATFTVDSNDEVVDDVVEYVNNKAKHGLAEVTD 843
OY 710 VDCV-PATTAKATASNVESVADTNVVAENGTVGAKGATSIYKNLTVPKD 759
DB 844 EKGMKDANIDKMYVGLVSDVRKDT-IYYKDADNKKKAS---IKSATVYFD 890

RESULT 12
PAC_STRMU STANDARD; PRT; 1565 AA.
AC P11657;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, last sequence update)
DT 01-OCT-1994 (Rel. 30, last annotation update)
DE PAC PROTEIN PRECURSOR.
GN PAC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OY NCBI_TextID=1309;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 39-48.
RC STRAIN=MT8148 SEROTYPE C;
RX MEDLINE=89343654; PubMed=2761390;
RA Okashiri N., Sasakawa C., Yoshikawa M., Hamada S., Koga T.;
RT "Molecular characterization of a surface protein antigen gene from

```

```

RT serotype C Streptococcus mutans, implicated in dental caries.";
RL Mol. Microbiol. 3:673-678(1989).
CC -1- FUNCTION: SURFACE PROTEIN ANTIGEN IMPLICATED IN DENTAL CARIES.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -1- SIMILARITY: STRONG TO AGGLUTININ RECEPTOR OF S.SANGUIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
CC -----
DR EMBL: X14490; CAA32652.1; -.
DR PIR: S04729; S04729.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00746; Gram_pos_anchor.1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING.1.
KW Antigen; Signal; Transmembrane; Repeat; Cell wall; Dental caries.
FT SIGNAL 1 38
FT CHAIN 39 1565
FT DOMAIN 39 1542 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1543 1560 MEMBRANE ANCHOR.
FT DOMAIN 1561 1565 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 219 464 3 X TANDEN REPEATS, ALA-RICH.
FT DOMAIN 1532 1537 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT PROTEINS.
SQ SEQUENCE 1565 AA; 170781 MW; 4C3B05C809D0C32A CRC64;

Query Match 5.2%; Score 201; DB 1; Length 1565;
Best Local Similarity 21.8%; Pred. No. 0.46;
Matches 177; Conservative 97; Mismatches 312; Indels 226; Gaps 37;

OY 25 GAVRGND-----KGMEPEKELTRAEATMAOILNLP-IDDKAKSFSDSQGWYTP 76
DB 53 GTGRCNATNLP-EAOGSASKAEBSQTLERQWHTIIVPTDLDQAKDKKSAG-----107
OY 77 FIAAVERAGVITGNGEFPNGKIDRVSMASLVEAYKLTDRNGTPATKREDETLNMG 136
DB 108 -VNVVQADADVKKGTVPKE-----EAVGKEIEIKEDYTKQAEIDIK-----147
OY 137 KERANIIVELGISGTGQWEPKKTIVTKAEAOFIATKDFGTEBAKVESAKAVTQKV 196
DB 148 -----TTDQYSDVAAHAEVAKIKAK-----NQATREOYEKMAAHKA 186
OY 197 EVKFSKAVEKLTG-----EDIKVTNRKN--NDKVLVKKEVLTSEDRSATVELYS 243
DB 187 EVERINNAAMASKAVEAKLQAYQADLAAYQKTAANOAAOKALAYQAELEKRVQE--A 244
OY 244 NLAAKQTY--TVDNKVKGTEVAVGSLEAKTIEMADQTVVADEPTAL-QFTVKDENGTEV 300
DB 245 NAAAKAAYDTAVANANNKNEIIAANEE---IRKRNTAKAEYETKLAQYGAELKRVQEA 301
OY 301 VSPGIEFVTPPAEKINA-KGEITFLAKGTSTVYAVYKKDKGVAAESKEVVSAGAAVA 359
DB 302 NAANEADY---QAKLTAYQTELARVOKANADAKATV-----EAAVAANNAKNA 346
OY 360 SISNMTVAEQNKADFTSKDFKONNKVYEGDNAYQVELKQDFNAVTTGKVEYE---SLN 415
DB 347 ALVAENATAIKQRENDNMAKATTEALIKQVEADLAIVK-----KANANEADYQAKLTAY 399
OY 416 TEVAVVDKATGKVTYLSAGAPVK--VTYKDSKSKALVSHTEIEA--FAQAKMKDIKLE 471
DB 400 TELARVOKAN-----ADAKAAVEAAVAANNAALTAENTAIKKRNADAKADYEAKLA 453
OY 472 K--TNVALSTKDVYDLKVKAPVLDQYKETEPAPTYAVVLDKDGKELKEQLKAYVREL 529
DB 454 KYADLAKTYOKDLADY-----PVKIKAYEDEQTSIKALAE-----489

```

QY 530 VLNAGQAGNTVTVLTKSGEK-----EAKATLAE-----LKAPG---AFSKFEVRGL 576
DB 490 -LEKHEKNEGDN---LTPESAOLVLYDEPNANLSLTGDKFKAKAVDAFSK----- 538
QY 577 DRELDKYVEENOKNAMTVSVLPVDANGVLKGAEEAEKLVTTNKEGKVDATDAQVTV 636
DB 539 STSKAKY-----DOK-----ILQD--DLDITYLEOSNDVASSMELKGNFGDAAGSTTV 586
QY 637 QNNS-----VTVGQAKAGETVTVVLDGKLIYTHSFK--VVDAPTAKGLAV---E 685
DB 587 SNNSQVKGMSVLLERGOSATATYTNLONSNGKISKIYKYTVDPKSKFGQKXVWLGI 646
QY 686 FVTSLSKEVAPNADLKALLNLISVDCVPAITTKAKATASNEFVSADTNYVAENGTVGAKG 745
DB 647 FTDPTL-----Gv-----FASAVTGOYEKN----- 666
QY 746 ATSIYVKNLTVYVDGKEQKVEPKAVOVAESI 777
DB 667 -TSIFIKNEFTFHEDEKPIEDNALLSVTSL 697

RESULT 13

GUN_BAC56 STANDARD: PRT: 941 AA.
AC P19424:
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ENDOLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (ALKALINE CELLULOSE)
OS Bacillus sp. (strain KSM-635).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxId=1415;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91037937; PubMed=230718;
RA Ozaki K., Shikata S., Kawai S., Ito S., Okamoto K.:
RT "Molecular cloning and nucleotide sequence of a gene for alkaline
RT cellulase from Bacillus sp. KSM-635.";
RL J. Gen. Microbiol. 136:1327-1334(1990).
CC -1 CATALYTIC ACTIVITY: ENDOLUCANASE OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1 SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -1 SIMILARITY: CONTAINS 3 S-LAYER HOMOLOGY (SLH) DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: M77420; AAA2304.1; -
CC PIR: S29043; S29043.
DR InterPro: IPR001347; Glyco_hydro_F5.
DR InterPro: IPR001119; SLH.
DR Pfam: PF00150; cellulase: 1.
DR Pfam: PF00395; SLH: 3.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
DR PROSITE: PS01072; SLH_DOMAIN: 2.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat.
FT SIGNAL 1 29
FT CHAIN 30 941 ENDOLUCANASE.
FT DOMAIN 40 99 SLH 1.
FT DOMAIN 100 151 SLH 2.
FT DOMAIN 152 225 SLH 3.
FT ACT_SITE 373 373 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 485 485 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 941 AA; 104628 MW; BEA2AC3B169BRADA CRC64;

Query Match 5.1%; Score 199; DB 1; Length 941;
Best Local Similarity 20.2%; Pred. No. 0.31;
Matches 179; Conservative 126; Mismatches 343; Indels 238; Gaps 42;

QY 5 FPDVPADHHGIDSINVLVEKGAIVKNDKGMFEGRKLTREAEATMAQILNLPIDDAKP 64
DB 41 FSDVKRTSFPYIKDLYDEVTGTSATTFSPDSTVTRQFTVTLRLGLGLEASSKDPY 100
QY 65 SFADSGOWTPPTIAAVERKAGVIKGTGNG--FEPNGKIDRSMASLLVEAYK--LDTRVKN-G 121
DB 101 -FKDRK--MNAYPEKIQAAIYERGIYTGKTNGEFAPNENITRQMAAAMAVRAYETYLENLSLP 158
QY 122 TPATKREDELTLNW-----GKEKANIIVELGISVGTDOMEPKKTVTYKAAQFI----- 171
DB 159 EEOREYNDSSSISTFAQDANQKAYVL-EL-MEGNTDGYFQPKNSTREQSAKIYSLTLMK 216
QY 172 -AKTDKQFGFEAAKVASAKAVTTQKVEV--KSKAVYEKLKEIK--VINKANNDKVLVK 226
DB 217 VASHDLYLHTEAVKSPS--EAGALQVELNGQLTLAEGEDGTPVOLRGMSTHGLQWFEIYN 275
QY 227 E--VTISEDKRSATVELYSNLAAKQTYTVDVNKGV--KTEVAVGSLEAKTTEMA---DOT 279
DB 276 ENAFVALSNDMGSMTRL-----AMYGENTATNPPEVKDLYEGIELAFEHDMY 325
QY 280 VVAD-----EPTALQFT-----V 292
DB 326 VIVDMVHVAQGRADRVYSGAYDFEEIADHYKDHPKNHYIWELANEPS--PNNNGPGILT 385
QY 293 KQDNQTEVVSPEIEFVTYPAEKINKGETTLAKGSTTYKAYKKDKGVVAASKEVYKS 352
DB 386 NDKGWEAYK---EYAEPLVEMLREKGDMLLVGN----- 417
QY 353 AEGAAVASISNMTVAEONKAD--FTSKDFKONKKVEGDNAVYOV-----ELKQDFNAV 404
DB 418 -----PMSQORPLSADNPIDANENIMYSVHFTGSGASHIGYPSGTPSESSSNM 468
QY 405 TTGKVEYESINTEVAVVDKATGVTVLSAGKAPVYTVKDSKKALVSHTVETLEAFQAQA 464
DB 469 --ANVRV--ALDNGVAVF--ATEWGTQANGDGGPYFDEADVWNLFLNKHNTSWANNSLTN 523
QY 465 MKDKLEKTNVALSTKDVNDLVKAPVLYOYKEFPAVTVYKLDGKRELKQKLEAYK 524
DB 524 KNEISGATFPFELGRTDAD-----LDPGANQVAP-----BELSLSGEY 563
QY 525 VNRLEVLNAGQAGNTVTVLTKSGEKEKATLAEKAPAFSKFEVRGLDTELDKYV 584
DB 564 VRARI-----KGIEYTPI-----DRTKFKLVWDFNDGTGQFQVNG--DSPKKEI 608
QY 585 TEENQKNAMTVSVLPVDANGVLKGA--EAEELKVTTNKEGKVDATDAQVTV-----Q 637
DB 609 TLTNNNDALDIEELVNSND--ISEGYNWNVRLSADGWSENDVILGATELTIDIVIEEPT 666
QY 638 NNSVITVGGQAKA-----GETYKRVTVLDGKLIYTHSFKVVDTAPT 678
DB 667 TVSIAAIPOGPAAGMANPTRAIVTEDEDFSGDGKALKVTTISE--DSLSLETIATSP 724
QY 679 AKGLA--VEFTSTLSKEVAPNADLKALLNLISVDCVPAITTKAKATASNEFVSADTNYVA 736
DB 725 DNTMSNLIIFVGT-----DAD-----VISLDNI--TVSGTEIEIIVHDE----- 763
QY 737 ENSTVAGKATSIYVKNLTVYVDGKEQKVEF--DKAVOVAVSIKEA 780
DB 764 -----KGTATL-----PSTFEDGTROGMDWHTESGVKTALTEEA 798

RESULT 14

NCAL_CHICK STANDARD: PRT: 1091 AA.
ID NCAL_CHICK
AC P13590; Q90919; Q90918;
DT 01-JAN-1990 (Rel. 13, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE NEURAL CELL ADHESION MOLECULE, LARGE ISOFORM PRECURSOR (N-CAM 180)
 DE [CONTAINS: N-CAM 140].
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauromorpha; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 DX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE OF 1-175 FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=87206190; PubMed=3576199.
 RA Cunningham B.A., Hemperly J.J., Murray B.A., Prediger E.A.,
 RA Brackenbury R., Edelman G.M.;
 RT "Neural cell adhesion molecule: structure, immunoglobulin-like
 RT domains, cell surface modulation, and alternative RNA splicing.";
 RL Science 236:799-806(1987).
 RN [2]
 RP SEQUENCE OF 128-1091 FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=86206089; PubMed=3458261.
 RA Hemperly J.J., Murray B.A., Edelman G.M., Cunningham B.A.;
 RT "Sequence of a cDNA clone encoding the polysialic acid-rich and
 RT cytoplasmic domains of the neural cell adhesion molecule N-CAM.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:3037-3041(1986).
 RN [3]
 RP SEQUENCE FROM N.A. (726 AA FORM).
 RX MEDLINE=87092340; PubMed=3467341.
 RA Hemperly J.J., Edelman G.M., Cunningham B.A.;
 RT "cDNA clones of the neural cell adhesion molecule (N-CAM) lacking a
 RT membrane-spanning region consistent with evidence for membrane
 RT attachment via a phosphatidylinositol intermediate.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:9822-9826(1986).
 RN [4]
 RP SEQUENCE OF 810-1069 FROM N.A.
 RX MEDLINE=87033934; PubMed=3771645;
 RA Murray B.A., Owens G.C., Prediger E.A., Crossin K.L.,
 RA Cunningham B.A., Edelman G.M.;
 RT "Cell surface modulation of the neural cell adhesion molecule
 RT resulting from alternative mRNA splicing in a tissue-specific
 RT developmental sequence.";
 RL J. Cell Biol. 103:1431-1439(1986).
 RN [5]
 RP SEQUENCE OF 1-17 FROM N.A.
 RX MEDLINE=93122797; PubMed=1478668;
 RA Colwell G., Li B., Forrest D., Brackenbury R.;
 RT "Conserved regulatory elements in the promoter region of the N-CAM
 RT gene.";
 RL Genomics 14:875-882(1992).
 RN [6]
 RP SEQUENCE OF 1-17 FROM N.A.
 RC STRAIN-WHITE LEIGHORN; TISSUE=Erythrocyte;
 RA Sasner M., Covault J.;
 RL Submitted (AUG-1993) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
 CC NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
 CC NEURITES, ETC.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF
 CC N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M21178; AAB59958.1; JOINED.
 DR EMBL: M21179; AAB59958.1; JOINED.
 DR EMBL: M21180; AAB59958.1; JOINED.
 DR EMBL: M15929; AAB59958.1; JOINED.
 DR EMBL: M15930; AAB59958.1; JOINED.
 DR EMBL: M15931; AAB59958.1; JOINED.
 DR EMBL: M15932; AAB59958.1; JOINED.
 DR EMBL: M15933; AAB59958.1; JOINED.
 DR EMBL: M15934; AAB59958.1; JOINED.
 DR EMBL: L29437; AAB59958.1; JOINED.
 DR EMBL: M15935; AAB59958.1; JOINED.
 DR EMBL: M15937; AAB59958.1; JOINED.
 DR EMBL: M15938; AAB59958.1; JOINED.
 DR EMBL: M15939; AAB59958.1; JOINED.
 DR EMBL: M15861; AAB59958.1; JOINED.
 DR EMBL: M15860; AAB59959.1; JOINED.
 DR EMBL: M15922; AAB59959.1; JOINED.
 DR EMBL: M15923; AAB59959.1; JOINED.
 DR EMBL: M15924; AAB59959.1; JOINED.
 DR EMBL: M21178; AAB59959.1; JOINED.
 DR EMBL: M21179; AAB59959.1; JOINED.
 DR EMBL: M21180; AAB59959.1; JOINED.
 DR EMBL: M15929; AAB59959.1; JOINED.
 DR EMBL: M15930; AAB59959.1; JOINED.
 DR EMBL: M15931; AAB59959.1; JOINED.
 DR EMBL: M15932; AAB59959.1; JOINED.
 DR EMBL: M15933; AAB59959.1; JOINED.
 DR EMBL: L29437; AAB59959.1; JOINED.
 DR EMBL: M15935; AAB59959.1; JOINED.
 DR EMBL: M15937; AAB59959.1; JOINED.
 DR EMBL: M15939; AAB59959.1; JOINED.
 DR EMBL: M15936; AAB59957.1; JOINED.
 DR EMBL: M15860; AAB59957.1; JOINED.
 DR EMBL: M15922; AAB59957.1; JOINED.
 DR EMBL: M15923; AAB59957.1; JOINED.
 DR EMBL: M15924; AAB59957.1; JOINED.
 DR EMBL: M21178; AAB59957.1; JOINED.
 DR EMBL: M21179; AAB59957.1; JOINED.
 DR EMBL: M21180; AAB59957.1; JOINED.
 DR EMBL: M15929; AAB59957.1; JOINED.
 DR EMBL: M15930; AAB59957.1; JOINED.
 DR EMBL: M15931; AAB59957.1; JOINED.
 DR EMBL: M15932; AAB59957.1; JOINED.
 DR EMBL: M15934; AAB59957.1; JOINED.
 DR EMBL: L29437; AAB59957.1; JOINED.
 DR EMBL: M15935; AAB59957.1; JOINED.
 DR EMBL: M15937; AAB59957.1; JOINED.
 DR EMBL: X04479; CAB51638.1; JOINED.
 DR EMBL: X70342; CAA49807.1; JOINED.
 DR EMBL: Z12128; CAA78113.1; JOINED.
 DR PIR: A43613; ICHNL.
 DR InterPro: IPR001777; FN_III.
 DR InterPro: IPR003006; IG_MHC.
 DR InterPro: IPR003598; IG_C2.
 DR InterPro: IPR003600; IG_Like.
 DR Pfam: PF00041; tn3; 2.
 DR Pfam: PF00047; tn3; 5.
 DR SMART: SM00060; FN3; 2.
 DR SMART: SM00408; IGC2; 2.
 DR SMART: SM00410; IG_Like; 1.
 KW Cell adhesion; Glycoprotein; Transmembrane; Repeat;
 KW Immunoglobulin domain; Alternative splicing; Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 1091
 FT DOMAIN 20 711
 FT FT 712 729
 FT DOMAIN 730 1091
 FT FT 1091
 FT DOMAIN 34 103
 FT FT 103
 FT DOMAIN 132 196
 FT FT 196
 FT DOMAIN 228 294
 FT FT 294
 FT DOMAIN 322 392
 FT FT 392
 FT DOMAIN 419 486
 FT FT 486
 FT DOMAIN 518 595
 FT FT 595

NEURAL CELL ADHESION MOLECULE, LARGE
 ISOFORM.
 EXTRACELLULAR (POTENTIAL).
 POTENTIAL.
 CYTOPLASMIC (POTENTIAL).
 IG-LIKE C2-TYPE DOMAIN 1.
 IG-LIKE C2-TYPE DOMAIN 2.
 IG-LIKE C2-TYPE DOMAIN 3.
 IG-LIKE C2-TYPE DOMAIN 4.
 IG-LIKE C2-TYPE DOMAIN 5.
 FIBRONECTIN TYPE-III 1.

Oy	91	GNCEPGRKIDRYSMAALLEAVKLLDTKVNGPPTATKFXDLLEFLMNGKEKANLVBELGISV	150
Dp	345	AN-----DDV-YSTIIIPAIQIN-----	380
Oy	151	GTGDOMEPRKTVTKAEAAQFIAKTDKQFGEAKAYESAKAVTTOKVEYFSAKVELTKE	210
Dp	381	IAGVAMP--SSVRFFGQYDVVATTTAAVQGPYKGVVILAGVQNDQYQ--KVINEIYEN	438
Oy	211	DIKATNKANDKVLKVEVLTSEDKRSATVVELS-----	254
Dp	439	DOAVLKFEYLDPTPKGCEVDRGDTDTGFIELTLLSPYTKVGTIYDYNINSKITPAFTID	498
Oy	255	VNK-----	275
Dp	499	PTKRVNVPFKSEONEGSKRYIAQNSGDETTKGICYIATQVWTKLGTNMGWPDY	558
Oy	276	ADQTVNVADE-----PTALOFTVK-----	296
Dp	559	SD-----DQAGIKFNNKGFPMPAGVQNTLRNAPATAVETTYIYKSESKGDIYEXYDPTD	613
Oy	297	GTEVY-----SPECIEFVT-----PAA-----	329
Dp	614	GKOIVNSVVDTPKPSALCTEYNTDVRDRPASIVAAQDTVFYFEKVSQDSAKTGTGVAGT-	672
Oy	330	TTVKAAYVKKKGKVYA-----ESKEVK--VSABEAAVASISMTVAEQNKAD--FESKCP-	379
Dp	673	TTTVKVEYEEKGSVNVNFNDINGKVIKAPASDEKDAKPGVNYDPTDLDQKLASTIFTEGKRYK	732
Oy	380	-----KONKKVEYEGDNAYVOVELKDOFNAVTTGKVEYESLTEEVAVVDKA-T	425
Dp	733	LVPAQDVPYKGVKGNLLEVGNNTKAKGD-----PTTGKIE-AGVAKETTYYYRAVT	784
Oy	426	GKVTYV-----	445
Dp	785	GSVVVNVKRDTEGNVICKDPEDVSDAIVGDAVYTTDDKKPNEIITTKDGSRYVLVPSKTDGEE	844
Oy	446	KGRALVHTEYIEIAFNQKAMK-----	474
Dp	845	NGK-VIEGSTITVYVYOKVANMIPETIPNVPEIDRPKVYRPEDPTPEDEPIDETPTGTCGE	903
Oy	475	-----VALSTKRDVTDLKAKAVLADQYGEFTAP-----	506
Dp	904	VPNIPYVPGTTPVDPDKONTPLKPIDP--NDPGKGYVPPRPENRPGVDTPIPYVPAKKVYTN	961
Oy	507	VLDKQKELKEQKLEAKYVRELVLVNAQDQAGNVTVLJTAASGEKEAKATLALCLKAPG	566
Dp	962	HVDEGGMPIAPOE-EGTKPRKSI-----PGYE-----FTGKTVTDEOGNTTHIYKTP-	1008
Oy	567	AFSAFEYRGDTELDKVYVEENQKNMNTVSLVPLVANGSLVGLGAEALAYVTTTKBEKE	626
Dp	1009	-----EKN-GTYVNVNVTVD-----GVYK-----EPDTPPTSPSEPT	1042
Oy	627	VDATDAOVTVQNSVLTIVGAGAKGETYKVT-----VLVDGKLITTHSFKVVDT---	675
Dp	1043	YDTJTD-----NKPRII--TFKGGEYIELVRDGTGENKVVGEGETVYIYVYRKETPAK	1092
Oy	676	-----APTAKGL-----AVEFTSISLKE-----VAPNADLKALANI	707
Dp	1093	KVVTNVHVEDEGNPVAPOEBEGTKPNKSI PGEEFGTKVTDDEGNTTHIYKTPAKKVYTNH	1152
Oy	708	LSYDGVY-----ATPAKATASVNEVYSADTNTVNAENGTVGAKKATSIYK	752
Dp	1153	VDEGNGTIAQOEDGTPPKROIISGEYV--RTVVDGEG-----NTHHYK	1195

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2002, 07:51:27 ; Search time 111.59 Seconds
(Without alignments)
1028.979 Million cell updates/sec

Title: US-09-754-947-1
Perfect score: 3885
Sequence: 1 AGKTFPDVPADHWGIDISIN.....EPDKAVQAVSIEKAPATK 785

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 segs, 14627329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL.17 : *
1: sp_archaea : *
2: sp_bacteria : *
3: sp_fungi : *
4: sp_human : *
5: sp_invertebrate : *
6: sp_mammal : *
7: sp_mhc : *
8: sp_organelle : *
9: sp_phage : *
10: sp_plant : *
11: sp_rodent : *
12: sp_virus : *
13: sp_vertebrate : *
14: sp_unclassified : *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3304.5	85.1	816	2	Q9ZES5 bacillus th
2	770.5	19.8	821	2	Q9RED0 bacillus th
3	748.5	19.3	823	2	Q45759 bacillus th
4	719	18.5	862	2	P94217 bacillus an
5	487.5	12.5	531	2	Q9RMZ0 bacillus an
6	449.5	11.6	931	2	Q91655 bacillus fi
7	445	11.5	1099	2	Q68840 bacillus st
8	415	10.7	920	2	Q45664 bacillus st
9	398.5	10.3	1047	2	Q9RB35 cytophaga s
10	348.5	9.0	1268	2	Q9REB7 bacillus sp
11	279.5	7.2	404	2	Q9X324 bacillus an
12	268.5	6.9	969	5	Q9ND19 plasmidium
13	268	6.9	1036	2	Q86999 clostridium
14	259	6.7	1109	2	Q53505 campylobact
15	249.5	6.4	3198	5	Q9UB88 manduca sex
16	248.5	6.4	1112	2	Q52781 campylobact
17	245.5	6.3	1822	2	Q07290 streptococc
18	241	6.2	1128	5	Q26947 trypanosoma
19	240.5	6.2	2353	2	P71401 haemophilus

20	240	6.2	3488	5	P91257	P91257 caenorhabdi
21	237.5	6.1	652	2	Q9X360	Q9X360 bacillus an
22	235.5	6.1	1344	2	Q49545	Q49545 mycoplasma
23	234.5	6.0	1795	2	Q9LCJ9	Q9LCJ9 staphylococ
24	234.5	6.0	6642	5	Q01761	Q01761 caenorhabdi
25	232.5	6.0	1557	2	Q9RN12	Q9RN12 haemophilus
26	232.5	6.0	1758	2	Q9JMS5	Q9JMS5 escherichia
27	232.5	6.0	1829	5	Q22248	Q22248 caenorhabdi
28	231	5.9	903	2	Q9K105	Q9K105 bacillus st
29	231	5.9	1432	2	Q9F292	Q9F292 yersinia pe
30	231	5.9	2349	2	P94750	P94750 escherichia
31	231	5.9	2383	2	P76347	P76347 escherichia
32	230	5.9	2402	2	Q9AER7	Q9AER7 staphylococ
33	229	5.9	6632	5	Q17362	Q17362 caenorhabdi
34	228.5	5.9	2055	2	Q85472	Q85472 abiotrophia
35	227.5	5.9	5458	2	Q9U459	Q9U459 plasmidium
36	226	5.8	1097	2	Q9ZAI7	Q9ZAI7 anaerobic t
37	225.5	5.8	2045	2	Q9AOK5	Q9AOK5 streptococc
38	223.5	5.8	3241	2	Q9AHR9	Q9AHR9 fusobacteri
39	222	5.7	565	2	Q30921	Q30921 mycoplasma
40	221.5	5.7	6677	5	Q9N435	Q9N435 caenorhabdi
41	220.5	5.7	1148	2	Q60045	Q60045 thermoaer
42	220.5	5.7	2478	2	Q9RL69	Q9RL69 staphylococ
43	220	5.7	1752	2	Q9A852	Q9A852 ruminooccu
44	218.5	5.6	1302	2	Q49547	Q49547 mycoplasma
45	218.5	5.6	1323	2	Q9AL34	Q9AL34 clostridium

ALIGNMENTS

RESULT	ID	Q9ZES5	PRELIMINARY:	PRT:	816 AA.
AC	Q9ZES5	01-MAY-1999 (TREMBLrel. 10, Created)			
DT	01-MAY-1999 (TREMBLrel. 10, Last sequence update)				
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)				
DE	CTC PROTEIN.				
GN	CTC.				
OS	Bacillus thuringiensis.				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;				
OX	Bacillus/Staphylococcus group; Bacillus.				
RN	NCBI_TaxID=1428;				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CTC;				
RA	Sun M., Yu Z.;				
RT	"S-layer protein gene of Bacillus thuringiensis CTC."				
RL	Submitted (Oct-1998) to the EMBL/Genbank/DBJ databases.				
DR	EMBL; AJ012290; CAA09981.1; -				
DR	InterPro: IPR001119; SLH.				
DR	InterPro: IPR00343; Big_2.				
DR	Pfam; PF00395; SLH; 3.				
DR	Pfam; PF02368; Big_2; 1.				
RW	S-layer.				
SO	SEQUENCE	816 AA; 87293 MW; 1002892905802151 CRC6;			
Query Match	Best Local Similarity	85.1%; Score 3304.5; DB 2; Length 816;			
Matches	679; Conservative	33; Mismatches	68; Indels	5; Gaps	5;
Qy	1	AGKTFPDVPADHWGIDISINYLVEKAVGNDKGFPEPKELTRAAATMMAOILNLPDK			60
Db	30	AGKSPDVPADHWGIDISINYLVEKAVTGNDKGFPEPKELTRAAATMMAOILNLPDK			89
Qy	61	DAKPSFADSGOGWYPTFLAAVEKAGVIGKTGNGFEPNGKIDRVSA\$LLVEAYKLDTRVN			120
Db	90	DAKPSFADSGOGWYPTFLAAVEKAGVIGKTGNGFEPNGKIDRVSA\$LLVEAYKLDTRVN			149
Qy	121	GTPAKRFKDLFTLNMGRKANILVELGTSVG-TGQWMPKRTVYTA\$AOPFAKIDKDFG			179

```
Db 150 GTPATKFKDLETLNMGKEKANILVELGISVGTADKWEPRKVTYKAEOAQTAKTDKQFG 209
Qy 180 TEAAKYESAKAVTTQKVEVFKSAVEKLTREKEDIKYTNKANNDKVLKREYLTSEDKRSATY 239
Db 210 TEAAKYESAKAVTTQKVEVFKSAVEKLTREKEDIKYTNKANNDKVLKREYLTSEDKRSATY 269
Qy 240 ELYSNLAQKQTYTVDNKVGKTEVAVGSLEAKTIEMADQTVVADEPTALOFTYKDENGTE 299
Db 270 ELYSNLAQKQTYTVDNKVGKTEVAVGSLEAKTIEMADQTVVADEPTALOFTYKDENGTE 329
Qy 300 VVSPEIEEVTTPAAEKINAKGEITLAKGSTYTKAYKKDKGVAAASKVEKVAEAGAAVA 359
Db 330 VVSPEIEEVTTPAAEKINAKGEITLAKGSTYTKAYKKDKGVAAASKVEKVAEAGAAVA 369
Qy 360 SISNMVVAEONKADFTSKDFKQNNKYEGDNAYVQVELKDOFNAYTTG-KVEYESLNTVEY 418
Db 390 SISNMVVAE-KADFTSKDFKQNNKYEGDNAYVQVELKDOFNAYTTG-KVEYESLNTVEY 448
Qy 419 AVVDKATGKVTYVLSAGKAPVKTVDKSKGALVSHTEIEAFAQKAMKDIKLEKTVAIS 478
Db 449 AVVDKATGKVTYVLSAGKAPVKTVDKSKGALVSHTEIEAFAQKAMKDIKLEKTVAIS 508
Qy 479 TKDVTDLKAPVLDQYGEKFTAPVTYVLYDKDGEKLEKQLEAKYVNBELVINAQGEA 538
Db 509 TKDVTDLKAPVLDQYGEKFTAPVTYVLYDKDGEKLEKQLEAKYVNBELVINAQGEA 568
Qy 539 GNTVTVLTAKSGEKEKATLAELEKAGAFSKEFVEGLDTELDRKYTEENOKNMTVSVL 598
Db 569 GKTYVELTAKSGEKEKATLAELEKAGAFSKEFVEGLDTELDRKYTEENOKNMTVSVL 628
Qy 599 PVDANGLVLRKGAFAELKATTTTKEGKEDATDAQTVYVONNS-VITVGGAKAGETTYK 657
Db 629 PVDANGLVLRKGAFAELKATTTTKEGKEDATDAQTVYVONNS-VITVGGAKAGETTYK 688
Qy 658 VLDGKLITTHSEKVVDTAPTAAGLAVETSTLSEVAPNADLKAALLNLISVDGPAT 717
Db 689 VLDGKLITTHSEKVVDTAPTAAGLAVETSTLSEVAPNADLKAALLNLISVDGPAT 748
Qy 718 AKATASNEFEVSADTNVA-ENGTVGAKGATSIYKVLTVVKGDKGQKVEFDKAVOAVS 776
Db 749 AGATVDFVDFVSADTNVSEETAKFGTKGSTIFVKELFVKKGEOQKVELDKPRVADVS 808
Qy 777 IREAK 781
Db 809 IREK 813

RESULT 2
Q9REDO PRELIMINARY: PRT: 821 AA.
AC Q9REDO: 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE SURFACE-LAYER PROTEIN PRECURSOR.
CN SLPA.
OS Bacillus thuringiensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus.
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1428;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 4045;
RA Mesnage S., Hausant M., Gounon P., Fouet A.;
RT "Characterization of B. thuringiensis S-layer.";
RL Submitted (Sep-1999) to the EMBL/Genbank/DDbJ databases.
DR EMBL: AJ249446; CAB63252.1;
DR InterPro: IPR001119; SLH.
DR Pfam: PF00395; SLH; 3.
KW Signal.
FT SIGNAL. 1 29
SQ SEQUENCE 821 AA: 87279 MW: 8068995C812214B8 CRC64:
```

```
Query Match 19.8%; Score 770.5; Db 2; Length 821;
Best Local Similarity 32.5%; Pred. No. 2.8e-26;
Matches 283; Conservative 114; Mismatches 305; Indels 169; Gaps 40;

Qy 1 AGTFFPDVADHWGIDISINLYVEKGAVKGNKQMEPEKELTPAEATYMAOQLNLPDK 60
Db 30 AGTFFPDVADHWGIDISINLYVEKGAVKGNKQMEPEKELTPAEATYMAOQLNLPDK 89
Qy 61 DAKPSFADSGOGWYPPFIAVEKAGVTKGTGNG-PEPNKIDRVASASLVEAKYKDFV 119
Db 90 GAKPSFKADQDSMAKTYIAVEQAGVIGDGTGNFSPQINASHASMTVAIKLEDKY 149
Qy 120 NGTPATKFKDLETLNMGKEKANILVELGISVGTQDQMEPRKVTYKAEOAQTAKTDKQFG 179
Db 150 SGLETFNFDLKD-HMGEDANILVALGTINGNMGMEPDKSTYRAEAKFIKTQMDQFG 208
Qy 180 TEA-AKVESAKAVTTQKVEVFKSAVEKLTREKEDIKYTNKANNDKVLKREYLTSEDKRSAT 238
Db 209 OKAEAKVESIKEINAKIEVEKTEGKEDVTAANFAVE--GSKELDIEKVELSKDKKSAT 266
Qy 239 VELYSNLAQKQTYTVDNKVGKTEVAVGSLEAKTIEMADQTVVADE--PTALOFTYKDE 295
Db 267 ITLKNALVKNQAVAHVKD-----VKSVDGKDIPKALEVIFPNDVAPTYSTVSTPBG 319
Qy 296 NGTEVVSPE-----GIEFVTPAAEKINAKGEITLAKGSTTV----- 332
Db 320 NVKVFSEKLSKDAVTYVINGKEFTATPEE-----NTVLTADVASVKNGEAFNIYVT 374
Qy 333 -----KAYKKDKGVYA-ESKEKYS--AEGAVASISNMVVAEONKAD 374
Db 375 GAKDLVGNTEMVEGKATYKVEKDVTAPEKDKVELVDGVAITLEV--TSEELSAO- 430
Qy 375 TSXDFKQNNKYEG-----DNAYVOV--ELKDOFNAYTTGKVEY-----ESLNTVEYA 419
Db 431 GKTYVELTAKSGEKEKATLAELEKAGAFSKEFVEGLDTELDRKYTEENOKNMTVSVL 486
Qy 420 VVDKATGKVTYVLSAGKAPVKTVDKSKGALVSHTEIEAFAQKAMKDIKLEKTVAIS 479
Db 487 VGNKVTREKGVKRDVYAPMLVKVVAADENKA-ATFTFDKQEVTAQEG--KLRYINLDTGS 540
Qy 480 KDVTDLKAPVLDQYGEKFTAPVTYVLYDKDGEKLEKQLEAKYVNBELVINAQGEA 539
Db 541 KDVTREVAVASVED-----NKKAITLTFQEGNKYVAA--TKGFVADTPAGNES 587
Qy 540 NYT-----VLTAKSGEKEKATLAELEKAGAFSKEFVEGLDTELDRKYTEENOKNMT 594
Db 588 AFTKEKVEKVEKKEGKDEKAPATKVERVADSKTKEFY-----TFDKVEY--KGGGADS 640
Qy 595 VSVLPVDANGVLKGAFAELKATTTTKEGKEDATDAQTVYVONNSVITVGGAKAGETTYK 654
Db 641 AS-----NVANNYTLGAKLEEGTLIVNADGK-----SVTIELPETFTF--EKSETV 685
Qy 655 KTYTV-----LDG-KLITTHSF-KVVDTP-APTAKGLAVETSTLSKEV-----APADTK 701
Db 686 KETVANAVANKDGKMGITMLLVNVYDTKAPERK--SAKITTKDAKITTLTPESEAVND-- 741
Qy 702 AALLNLISVDGPATYAKA--TASNVEFY-----SADTNVAENGTGAKGATSIYKVN 753
Db 742 -ATDFEIIDNGVLLPATKADETAESKDVLLKVTAPADVNLATGVTYKAVEGKRVVLN 800
Qy 754 LTVVKGDKGQKVEFDKAVOAVSIAKAPAT 784
Db 801 TADTSTINKNKLAVF-----KPYT 818

RESULT 3
AC Q45759 PRELIMINARY: PRT: 823 AA.
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
```



```
QY 382 NNK -VYEGDNAVVO---VELKDQFNAVTTCCKVEYESLNTFVAVVDKATGKVTVLASAGKAP 437
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 432 NSGNLVVGEKASLNKLVIATJAGEDKVDVDPQISIKSSNH--GIISVNNVITAEAGGEAT 489
QY 438 VKATVVDKSKALVSHVTELEAFQAKMADIKLEKTNVA---LSTADVVD-----L 485
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 490 LTIKVD-----VTKDKVKVTSDSKRLVSXANPDKLOVONKTL 530
QY 486 KVAPVLDOYGEKFTAPVTV--KVLDKDGKLEKQKLEAKYVRELV---LNAAGQAG 539
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 531 PTVFVTTDQGDPPGANTAIKEVLPRTG--VVAEGLDVYTTDSGSLGTITIGTVGNDVG 589
QY 540 NYTVVLTAASGEKAAATLAEEL-KAPGAFSKREV-----RGDTELD----- 581
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 590 EGTVHF--QNGNGATLGSLVNVTEGNVAFKNEFLVSKVGQYQSPDKLIDLNVSTVEY 647
QY 582 ---KYTTEENQKAMTVSVLPVDANGLVLGATAELKVTTCCKEKEVDATQATVQN 638
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 648 QLSKTVSDRVYSDPENLEGEVESKMLAADAQIVGKKNVVTKGTPOKVD----- 697
QY 639 NSVITVGGAGKAGETFKVTVLDGKLITTHSFKVVDTPAK----GLAVEFTSTSLKE 693
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 698 --IHLKNGATACKA-TVEIVQETIAKSYNFKPVQTEFVEKKNIGTVLELEKSLND 754
QY 694 VAPNADL 700
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 755 IVKGINL 761
```

```
RESULT 5
O9RM20 PRELIMINARY: PRT: 531 AA.
AC O9RM20:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PXO2-42.
OS Bacillus anthracis.
OC Bacillus anthracis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_Taxid 1392;
RN [1]
RP SEQUENCE FROM N.A.
RA Oklahoma R.T., Cloud K., Hampton O., Hill K.K., Keim P., Lamke G.,
RA Kumano S., Manter D., Martinez Y., Svensson R., Tatum L.R.,
RA Brown A.E., Jackson P.J.;
RL Submitted (NOV-1989) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF188935; AAF13647.1; -.
DR InterPro: IPR001119; SLH.
DR InterPro: IPR002508; Amidase_3.
DR Pfam: PF00395; SLH. 3.
DR Pfam: PF01520; Amidase_3; 1.
KW plasmid.
SQ SEQUENCE 531 AA; 58864 MW; 9A171DC4ED05CA78 CRC64;
```

Query Match 12.5%; Score 487.5; DB 2: Length 531;
Best Local Similarity 28.4%; Pred. No. 4e-14;
Matches 160; Conservative 93; Mismatches 188; Indels 123; Gaps 20;

```
QY 3 KTFPDVADHWGIDISINYLVEKGAVGNKDMFEPGKELTRABAATMAQILMLPIDKDA 62
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 31 KTFDVP--NMAQSVNYLMKK--ALDGKPDGTFSPSEKIDRGSAAKLAMAVLGLQINKQA 87
QY 63 KPSFASQGWYTPFTAAVEKAGYIKGTGNG-FEPNGKIDRVSMASLLVAYLDPKRVNG 121
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 88 KPSFOAKNMHMASPYIAAYEKAGVIVGDGSGNPNPSKIDIDRASMASMLVAYVLMNNIIG 147
QY 122 TPATKFKDLETLNMGKEKANILVELGISVGTGQDWEPPKKTVTAKAAOPIAKTDKQFTE 181
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 148 DLPTQFEDLKG-HWGAKLANALVALGLISKTGDKMKPNGITRAEAVQPIAQTDMK---- 202
```

```
QY 182 AAKVESAKAVTTCQVEVEKESKAVERLTKEDIKVTNNKANDKVLKVEVTLSEDKRSATVEL 241
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 203 --KADTSKRMVNR-----HFITYHQPSLSSGVTSNQNHQPIIVVEKQADGMIKI 251
QY 242 YSNLAKQITYTVNVNKGKTEVAVGSLKAKTETMAQOYVADEPTALQOTVADENSTEV 301
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 252 VTNIGDKWPLYE---KRE-TIHSTFTYTPASHSSKVLGTHSPQTYVIEKKS--- 302
QY 302 SPGIEFVVPAAEKINAKGEITLAK-----GSTVTKAVYKKGKGVY 343
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 303 ---WIKRTNASQWLMDKNQTLTPKQNNFLEKTTIIDPGHGDIGGKHGIMNNSPVY 359
QY 344 AESKEVYSAEGAAVASISNMVTAEQNKADFTSKDKFRQNNKVEGDNAYVVELKQDFNA 403
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 360 YDT-AVRV-----QKLPAQKTPFTALTTRDAYS- 386
QY 404 VTTGKVEYESLNTFVAVVDKATGKVTVLASGAKAPVYTVYKDSGKALVSHVTELEAFQAK 463
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 387 -RPGKNATQSLKRVGEPAKKNKGDIF-----VSIHANGFNAGHGETEYKKAPOK 437
QY 464 AMKDI-----KLEKTNV-ALSTRD---VTDLVKAPVLDOYGEKFTAPTVAVL-- 508
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 438 SNPYVDSRLAKIKQKRLITALQTRDRCYKIGNIVL-----RENTMPSVLTELGR 489
QY 509 ---DKDGKEL-----KEQKLEAKY 524
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 490 VDNKADGKKLIDSPWQRARAEATY 513
```

```
RESULT 6
O9L655 PRELIMINARY: PRT: 931 AA.
AC O9L655:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE SURFACE LAYER PROTEIN.
GN SLPA.
OS Bacillus firmus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_Taxid=1399;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=OF4;
RC Kilmour R., Messner P., Guffanti A.A., Kent R., Kendrick N.,
RA Krulwich T.A.;
RT "Identification and functional characterization of a pH induced S-
RT layer protein in facultatively alkaliphilic Bacillus firmus OF4.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF242295; AAF68436.1; -.
DR InterPro: IPR001119; SLH.
DR Pfam: PF00395; SLH. 3.
SQ SEQUENCE 931 AA; 96855 MW; 6A9727171C0A78D0 CRC64;
```

Query Match 11.6%; Score 449.5; DB 2: Length 931;
Best Local Similarity 22.9%; Pred. No. 4.1e-12;
Matches 219; Conservative 120; Mismatches 310; Indels 309; Gaps 37;

```
QY 1 AGTTPDVADHWGIDISINYLVEKGAVGNKDMFEPGKELTRABAATMAQILMLPIDK 60
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 34 ADAKFSDVSSHMAIDIDINYLVEKGAIQGYPDCTFEPNNSITRAELAVLVANLTLGDVDS 93
QY 61 D-AKPSFASQG-QWTPFTAAV--EKAGYIKGTGNG-FEPNGKIDRVSMASLLVAYVLT 115
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 94 EYTTDKFSVDVPRATNMWNPYIAALVDDTEGVTDYENGTFEPNSWTITROEKAKWVAEYDI 153
QY 116 DTVNGSTPATKFKDLETLNMGKEKANILVELGISVG-TGQDWEPPKKTVTAKAAOPIAKT 174
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 154 EL-VEGKD-LPFTYDVGSL-WSTDIYINLASNGVAAGMTATTPAPRQREVILRAQYAAFIHRA 210
```

QY 175 DKQFGTEAKVE-----SAKAV-----TTQAV----- 196
Db 211 E-----VEEERIEVPGLVPTVSSVTAVNATTLTVLSDRTHEVTLDTALVANEATEVTFEE 266
QY 197 --EWEKSAVEKLTREDIKVTNKANNDDVLYKEVTLSEDDKRSATVELSNLAAKQTYTVD 254
Db 267 INDVEYTELVTVYELAVASVNAVNGKOLVNFOTOPVASSVLTSPASTISTVTGVLIS 326
QY 255 VNKVG-----KTEVAVGSLEAKTIEMADQTVVADEPTALQF-----TVKDENGTEVVSPE 304
Db 327 LNRSTDRHSNVDVGVAVERAVALSSDGKTLTVAPTGQFKNVDYTVVDASGQNTIPS 386
QY 305 GIEFVT-----PAEKI-----NAKEI----- 322
Db 387 VIKTISVDQTPAPIVNSVYVVPATDKFEVTLSEPIDSLTEGLRLINGQVPASGFDALTCGP 446
QY 323 -----TLAKGTSMT----- 331
Db 447 TNKLTFFARSSVATGTMATITTAGFSDAAGNFVTPSTTVPVTOOTTLALAVASLEQVNTQ 506
QY 332 --VKAVYKKDKVVAESKEKVSABGAAY-----ASISMTVAEONKADFTSK--DFKQNN 383
Db 507 KVLTFENKE--LNSASKTALESSTGVVYTRPNSGTSNTFAQNLVDNTGKTYDITLSD 564
QY 384 KYEBGDNAVY--QVELKQDFNAVTTGKVEYESLNTAEVAVDKATGKVTYLSAGKAPVKYT 441
Db 565 ATYTGNSEVFGITLTKAFDTVTNNK--NDLYSKSTILNKDVTVAFTVTSALASNRQA 621
QY 442 VDSKGAALVSHVTEIEAFQAKMDIKLEKTNTVALSTKDVTDLKVAKAVLDQYKEFPA 501
Db 622 I-----EYTLSCGVITTNPAQVKLRKDKAEQGTGLSVALKGGTD----- 659
QY 502 PVTAKVLD-KDGELEK-----EOKLEA-----KYVNBELY-----LNAQO 536
Db 650 --NVLVSVSSGAELAGSGYQVLEGAVTDLGNANNAVNAPSVSATPAAPLNVAYA 717
QY 537 EAGNTVVLITAKSGEKAATLALBKAGASKEFEVRGLDELKYYTEENKNAAYVS 596
Db 718 NSGTVFVTAJTGQTFETASL-----NNNFKIDQAVSSNSDITLNTSTRITIVS 769
QY 597 VLPVDA-----NGLVFKGAEMAEKLVTTNKEGKEVDADAQVTVONNSVITVG 645
Db 770 LPSEDSVKISGALFTTNCLAL-----ESRALATATATVYTDN----- 809
QY 646 OGAKAGETVYKTVLDGLKLTTHSEKVVDTAPATAGLAVEFTSTSLKEVAPNADKALL 705
Db 810 -----TAPTLGAGQL-----VSANYI 825
QY 706 NILSDGVPAITAKATASVNEFVSADTNVVAENGTVGAKGATSIYKMLTYVK--DGK 761
Db 826 KTFEDENIDALTLTDAAIDLIDIOLSNGTVAYNGGF--ASGATS--GGDTVVSVYDCK 879

RESULT 7
ID 068840 PRELIMINARY: PRT: 1099 AA.
AC 068840:
01-AUG-1998 (Tremblrel. 07, Created)
01-AUG-1998 (Tremblrel. 07, Last sequence update)
01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE SURFACE LAYER PROTEIN.
GN SBSC.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=142;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC12980;
RA Jarosch M., Egelseer E.M., Mattanovich D., Sleytr U.B., Sara M.;
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF055578; AAC12757.1; -
InterPro: IPR003433; Big_2.

DR Pfam: PF02368; Big_2; 3.
SQ SEQUENCE 1099 AA; 115394 MW; 7330F156334ABBA CRC64;

Query Match 11.5%; Score 445; DB 2; Length 1099;
Best Local Similarity 24.6%; Pred. No. 8.1e-12;
Matches 223; Conservative 136; Mismatches 331; Indels 218; Gaps 41;

QY 44 AEAATMAAQLN-----LPIDKDKPSFADSGQWYTFPIAV 81
Db 28 SQAATDVAITVVSQAQAKQKKAATYTSHTVTETGCPDPDKDYAAATNKKK--DAYANAIVAV 86
QY 82 EKAGVIKGTG-----NGEPPNGKIDRVSNASLVEAYKIDTVNG-----TPAT 125
Db 87 NKAGAKKDAYLADQAIETVYFKANPKSGEARVAT--YIDAYNATKLDKMRQELKAAV 145
QY 126 KFKDLETLN-----WQKE-----KNIIVELGIS 149
Db 146 DAKDLKKAEBELYHKISTYELKTRTVILDRVYGOSTRELLRSTFFKADAQALRRLITDITVA 205
QY 150 VGTGDQWEPKK--TVYKAEEA-----QPIAKTDKQFTE-----AAKVE 186
Db 206 MKARAQDAVAVKAGNLDKAKKAAALDQVNOYVSKVTDFAFKAELOKAAQDAKATAEALTPKVE 265
QY 187 SAKAVTTQKVEYKESKAVEKLT--KEDIKVTNKANNDDVLYKEVTLSDKNSATVELYSN 244
Db 266 SVSAIDSTSFVYTFPKPYDKATAIKPNFSITLKGTEFLKRSVSESGQLATVTLYDT 325
QY 245 LAAKQTVVDVKKVKTVAVGSLEAKTIEMADQTVVADEPTALQF----- 290
Db 326 LYDGKTYIVVTS--GLKPTAGKEFETSTNEFLYKPV--PASITFNKMLPESAVULT 380
QY 291 --TVKDENGTEVVSPEGIEFTVPAER-----INAKGEITL-----AKGTS--TVKAV 335
Db 381 KVVYTKDAAGNVTKS--GELEFETSSEKLTQCKPINTGKKSIVINATVKGNTVTTGAVI 438
QY 336 YKKDQVVAESKEKVASBGAAVASISMTVAEONKADFTSKDFQONKKYVEGDNAVYQV 395
Db 439 LAVEDEKAAEVELKLTQDNKEVYTL--YANGNAFD-----KDGKQISG--TLTLTLA 487
QY 396 ELKQDFNAVTTGKV-----EYESLNTAEVAVDKATGKTVLSAGKAPKTVYKDSKGAL 450
Db 488 KFKDQYGMELTGKVACTDYTFESLNEPLVV--APGCVTPYIPGALVKKYKG-- 540
QY 451 VSHVTEIEAFQAKMDIKLEKTNTVALSTKDVTDLKVAPVLDQYKGEFTAPVTVKLDK 510
Db 541 VTKTIPVTVKANPVLJETIADVSTGVSAKGQKATPKV--TLKDQYGNKFTGVNV----- 593
QY 511 DGELEKQKLEAKYNNRELVLNAGQEGAGNTV-----VITAKSGEKAATIAL 560
Db 594 ----TSDKTETATVS-----VSNSGIGOSEYTVTVNGVAEGSTTITIKSGTREVK--VPV 642
QY 561 ELKAPGAFSKPEVRGLDT--ELDKVYTEENOKNAMVSVLPVDPANGLVKGAEALKVT 619
Db 643 NVVAGGPVANTQIKVLDGCKIDKSTFESPANNDDVLLKYYAANDANKNIY--GDTTNDVYTTIS 701
QY 620 --TNKEGKEVDATDAQVTVONNSVITVGOGAKAGETVYVTVLDGKLTTHSEKVVDTAP 677
Db 702 EATDINGVIYVANSKSTANGDPTVYITDNGSKYKGE--TLTYKLGIVTLGTVDVEVIDT-- 758
QY 678 TAKGLAVEFTSTSLK--EVAAPNADLKAALLNITLSV--DGVPAITAKATASVNE----- 726
Db 759 TLKATVTVTKKADLIELDAADNGDALAKLLANDIDQGNPNVDSAAAPNTNEKLOALKS 818
QY 727 ---FVSADTNVVAENGTV--GAKGATSIYKMLTVVKKQKEQV-----EPIKAVOVAISI 777
Db 819 VLSGIVSSDTVYIGSVSNVDNLKDDAST--SGLAVKKAAGTVTLTVFNEDSKIAPIALIV 876
QY 778 KEAKPATK 785
Db 877 K-APATQ 883

```

RESULT      8
ID 045664 PRELIMINARY: PRT: 920 AA.
AC 045664:
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE SRSB GENE (SEQUENCE 5 FROM PATENT WO9906567 PRECURSOR).
GN SRSB.
OS Bacillus steatothermophilus.
OC Bacteria: Firmicutes; Bacillus/Clostridium group:
OC Bacillus/Staphylococcus group: Bacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PV72;
RA MEDLINE=97197531; PubMed=9045827;
RA Kuen B., Koch A., Asenbauer E., Sara M., Lubitz W.;
RT Molecular characterization of the Bacillus steatothermophilus PV72 S-
RT layer gene srsB induced by oxidative stress.;
RL J. Bacteriol. 179:1664-1670(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=PV72;
RA Lubitz W., Resch S.;
RT "SECRETION OF CARRIER-BONDED PROTEINS INTO THE PERIPLASMA AND THE
RT EXTRACELLULAR SPACE.";
RL Patent number WO9906567-A/5, 11-FEB-1999. LUBITZ WERNER (AT);
RL RESCH STEPHANIE (DE);
DR EMBL: X98095; CA66724.1; -.
DR EMBL: AX000222; CAB7070.1; -.
DR InterPro: IPR003343; B1g_2.
DR InterPro: IPR001119; SLH.
DR Pfam: PF02368; B1g_2; 2.
DR Pfam: PF00395; SLH; 3.
KW S-layer; Signal.
FT CHAIN 32 920
SQ SEQUENCE 920 AA; 97916 MW; 1F3C20344B40F3A2 CRC64;

```

```

Query Match      10.7%; Score 415; DB 2: Length 920;
Best Local Similarity 24.4%; Pred. No. 1.3e-10;
Matches 207; Conservative 123; Mismatches 352; Indels 168; Gaps 35;

QY 4 TEPDVPADHMGIDSINYLVEKGAVGKNDKGMPEPKELTRAEATVMQAIIPLPIDKDAK 63
DB 33 SFTDVAPOY--KDAIDFLVSTGATKGTETKTCYDEITRLDAAYILAVLGLVDNMD 90
QY 64 PSFADSGOMYTFPIAAVEKAVIKGTGNG-FEPNGKIDRVSMASLVEAYKLD----- 116
DB 91 AGTDPKPD-RAKYVALVEAGVLNCKAPKGFAYDPLTRVEMAKTIANRYKLKADVKL 149
QY 117 --TKVNGTPATKFKDLETLNMGKERNLIVELGISVGTGDQDQEPKKTITKAEAFIKT 174
DB 150 PFTDVNDTMAPYKALYKYEYTKGK-----TPTSGAYONTIRGDFQAFVYRA 197
QY 175 DNOFGTEAKVSAVAVTQKVEKFSKAVEKLTREKIDIKVANKNDKVLVEKVTLSK 234
DB 198 VNI--NAVPEIETVAVNSTYKVFNTQI-----ADVDFNFAIDNGLVTKATLSK 250
QY 235 RSATVELYSNLAKOTYTVDVNKG--KTEVA-----VGSLEAKTIEADQTVVADE 284
DB 251 KSEVEVVKPFTRNQDITITATGINKLGETAKELTGKFWVSVDADVAVVALNNSSLK 310
QY 285 PIALQFTYKKGNGTEV-----SPGIEFTVPAEKINAKGEITLAKGTSTTVKAY 336
DB 311 ESSL--TVKDDGDKVDAKVELTSSNTNIYVSSGEVSAAAVTAVKGTADVTAKVT 368
QY 337 KKDGVVAES-----KEYKVASGAAVAASISNMTVAEO-----NKADFTSKDFKNNK 384
DB 369 LPDGVVLNTRKVVYTEPVVOVNOGFTLVNLSNAPONTVAFAKAKAVISMFGERTKV 428
QY 385 -VYE--GDNAVVOVELKDPQFNATVTGKVEYESLNTTEVAVVDKATGCVTVLS----AGKA 436

```

```

DB 429 AMYDTKNGDETRKPVDEK-----ATVRLNLIATAINGSELTVTYANAGOSGKA 479
QY 437 PVKTVVDSGKALVSHITVEIEFAQKAMKDIKLEKTNALVS-----TRQVTDLKVK 488
DB 480 SFEYTEKDN-----YKRTFTVDKPKDVLQDIDVATSVKLSDEAVGGEVGVNQKTIK 534
QY 489 APYLDQYKEF-----TAPVTYKVLDRD-----GKEIKQ----- 518
DB 535 VSAVDQKGEIKRGTGKVTYTTNTGLVIKVNNSNTIDPFGSADQEPVVAATKDKI 594
QY 519 ---KLEAKYVNRVLVNAAGQEAQNTVVLTAKSCEKAKAT-LALELKAPGAFSKFEVR 574
DB 595 VNGKVEKRYFKN-----ASDTPPTSTKRTITVNVVAVKADATPVGLDIAP--SEIDVN 645
QY 575 GLDTELDKRYTEENQKNAATVSLPVDANGVLKKGAEALKVTYTNKEGKEVDATDAQ 634
DB 646 APVT--ASTADVDFINFSEIYTLDSNGNKL-----KVPT--ATVLGTNDYV 692
QY 635 TYQNNSVITVGGG-----AKAGETVYVYVLDGLITTH-SFRVVDAPTAKGLAVEFTST 689
DB 693 EVNGNVLPQKGNDELTLTSSSTVNVVDATDG--ITKRIPVKYINSASVPATVATSPV 750
QY 690 SLEKVAPNADLKAALNLTLSVDGVPATTAKTATASVEEVSADTNVAEMGTGAKGATSI 749
DB 751 TVKLNSDDMLTFEELITGVID--PTQLVKDEIDIN-ETIAV-----SKAAKNDGYL 798
QY 750 YKKNLTIVKD 759
DB 799 YKKNPLVTAKD 808

RESULT      9
QY 09RB35 PRELIMINARY: PRT: 1047 AA.
AC 09RB35;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE S-LAYER PROTEIN PRECURSOR.
OS Cytophaga sp. 'Jeang 1995'.
OC Bacteria; CF8 group; Cytophagales; Cytophagaceae; Cytophaga.
OX NCBI_TaxID=85173;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JEANG 1995;
RX MEDLINE=95291197; PubMed=7539663;
RA Jeang C.L., Lee Y.H., Chang L.W.;
RT "Purification and characterization of a raw-starch digesting amylase
RT from a soil bacterium--Cytophaga sp.";
RL Biochem. Mol. Biol. Int. 35:549-557(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=JEANG 1995;
RA Jeang C.L., Liao T.W., Chiu S.Y., Kang P.L., Shieh T.Y.;
RL Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF068060; AAF21259.1; -.
DR InterPro: IPR000515; BPD_transp.
DR InterPro: IPR001119; SLH.
DR InterPro: IPR003343; B1g_2.
DR Pfam: PF00395; SLH; 3.
DR Pfam: PF02368; B1g_2; 2.
DR PROSITE: PS00402; BPD_TRANS_P_NN_MEMBER; UNKNOWN_1.
KW Signal.
FT SIGNAL 1 32 POTENTIAL.
FT CHAIN 33 1047 S-LAYER PROTEIN.
SQ SEQUENCE 1047 AA; 108718 MW; 4395402E9965295A CRC64;

```

```

Query Match      10.3%; Score 398.5; DB 2: Length 1047;
Best Local Similarity 24.2%; Pred. No. 8.4e-10;
Matches 228; Conservative 111; Mismatches 365; Indels 237; Gaps 40;

```


[illegible]

```

Oy 589 OKNMATYVLPVDMANGLYLKAEAAELKVTYTTNECK-----EVDATDAO-VYV 636
Db 704 -----VGTGDI-----ANAFALTNOSGTTLTNLGKMTYELRNTSGOPIEV 745
Oy 637 -----ONNSVLTVOGAGACETYKAVTVLDOGLKLTTHSFYKVDPAIPAKLAVEFTS 688
Db 746 KTIPIYRQNNWSSL---FRIEYGSITTSAD-VISTNAATIEIAPRGNKATVYVP 801
Oy 689 TSLKEVAPNADLKA-ALLNLTISVDC-----VPATTA-----KATASN---VEEVA 730
Db 802 SFV--TANNEPASADNNLNGVYVTDARDPFNKHVTPALPATAFEGNKKEVGTNYTGYOVFNK 859
Oy 731 DTNYVAENGST-----YKAKGATSIY 750
Db 860 DKKTITENGKNAVRYAGEGKTYRY 884

```

```

RESULT      11
09X324      ID      09X324      PRELIMINARY;      PRT;      404 AA.
AC          09X324;
DT          01-NOV-1999 (TREMblrel. 12, Created)
DT          01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT          01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE          PX01-54.
OS          Bacillus anthracis.
OC          Plasmid virulence plasmid PX01.
OC          Bacteria; Firmicutes; Bacillus/Clostridium group;
OC          Bacillus/Staphylococcus group; Bacillus.
OX          NCBI_TaxID=1392;
           [1]
RP          SEQUENCE FROM N.A.
RC          STRAIN-STERNF.
RA          Okinaka R.T., Cloud K., Hamton O., Hoffmaster A., Hill K.K., Keim P.
RA          Koehler T., Lamte G., Kumano S., Mahillon J., Manter D., Martinez Y.
RA          Rieke D.O., Svensson R., Jackson P.J.;
RT          "The sequence and organization of pX01, the large Bacillus anthracis
RT          plasmid harboring the Anthrax toxin genes.";
RL          J. Bacteriol. 0:0-0(1999).
DR          EMBL; AF065404; AACD3358.1; -.
DR          InterPro: IPR001119; SLH.
DR          Pfam: PF00395; SLH; 3.
DR          Prosite; PS01072; SLH_DOMAIN; UNKNOWN_1.
KW          Plasmid.
SQ          SEQUENCE      404 AA;      45044 MW;      45D0BFAA50C0C4C CRC64;

```

```

Query March 7.2% Score 279.5; DB 2; Length 404;
Best Local Similarity 25.8%; Pred. No.3.7e-05;
Matches 107; Conservative 66; Mismatches 161; Indels 79; Gaps

Oy 5 PPVADAHGIDISINYLVKCAVKGNDKGFEPKELTLRAAATMMAQIILPIDKA-- 62
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 48 FNDVADANIMSTKALYGLANRRVVGAYGNGOFGFEDDNRQVARMIAAYVK--PADADAF 106
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 63 KPSFVDSGGQMYPTPIAAVEKAGVYIKGTGNG--PEPNCKIDRVASMSILVAYKLYDKRV 121
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 107 KNPSSDIKGHMFEEKILALAEGLVAAYGECKKGPDDILTRQDAQVOLTNAEF--KA 162
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 122 TPATKFKDLEPLNMGCKEKANILVELGISVGTGDO--WEPRKKTVYKAEADQIAKTDKOF 180
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 163 TKTSFTDIDKNSMALKAISALENGVTIGIGDKLSPYAHVIREQYSGQLFNS----- 216
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 181 EAAVESKAKANTTOKVEKYSKAAVEKLTKEDIKVTNKAN--NDKVLKVELYLSKD--KRS 236
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 217 ---INVLKETKTEKVEPRKTRPETKPTKPTKPTKPTKPDVLDLKLVSNEFTPEKWEYESS 272
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 237 ATVELYSILAKQFYTVADNVKGTVEAVASGLEAKTITEMADQFVVADEPALQFTVKDEN 296
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 273 KVINSSSICQ--QSLINEVKKKYYTNLKYAEV--QGVIRLVD-----N 312
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 297 GTEVVSPEGI---EFTVPAAE--KI-----MAKELITLAKGSTVYKAVYKKDKGV--V 343
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

Db      313  G--MFLPEGSLKAQFIYIPINESGFKYFLDNDNDAIVELAKWTTLLKSLDNLDEIQGT 370
QY      344  AESKEVKVKSAGCAVAASISNNTVAEQKADPTSKDFQNNKVVYEGDAAYQVEIK 398
        || :: -|| -|| -|| -|| -|| -|| -|| -|| -|| -|| -|| -|| -|| -||
Db      371  VESHLNLTFFEGGNKVKVRIGNST-----ADNM-----MYQVSEK 404

```

RESULT 12		
ID	Q9ND19	PRELIMINARY; PRT; 969 AA.
AC	Q9ND19:	
DT	01-OCT-2000 (TREMBLrel. 15, Created)	
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)	
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)	
DE	MEROZOITE SURFACE PROTEIN 3G.	
GN	MSP-3G.	
OS	Plasmodium vivax (strain Belm).	
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
OX	NCBI_TaxID=31273;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=BELEM;	
RA	Gainski M.R., Al-Khedery B., Ingravallo P., Corredor-Medina C.,	
RA	Barnewell J.W.;	
RT	"Plasmodium vivax merozoite surface proteins-3b and -3g share	
RT	structural similarities with Plasmodium vivax merozoite surface	
RT	protein-3a and define a new gene family in Plasmodium.",	
RL	Submitted (OCT-1998) to the EMBL/Genbank/DBJ databases.	
SO	EMBL: AF099663; AAF8288.1;	
SR	SEQUENCE 969 AA; 104674 MW; 9A150CCA9918FF77 CRC64;	

	Query Match	6.9%	Score 268.5	DB 5:	Length 969;	
	Best Local Similarity	22.8%	Pred. No. 0.000387			
	Matches 185; Conservative	102;	Mismatches 335;	Indels 191;	Gaps 32;	
OY	23 EKGAYKGNDRKFPEEGKELTTRAEEA-----TMAAQILNLPIDDAKPSFASOGOWTPETI	78				
	: : : : : : : : : : : : : : : : : : :					
Db	110 EEDAKAKOR -LLEAVDKVKKAEELIKKGTAKAEALIKNAVNK-AK-EVADAKAKELTT---	163				
OY	79 AAVEAGAIYIKGTGNGFEENGKIDRVSMSLSLVEAYKKLDTKVNGTPATPKFDELTLMWKE	138				
	: : : : : : : : : : : : : : : : : : :					
Db	164 ---EKAKIAKEAKEBETAKIGKSVASFPALLIAEKAETEAK-----ANE	204				
OY	139 KANILIVEIGISVGDDOMEPPKTVTKAAEOPIAK-----TDKOFGPTA	182				
	: : : : : : : : : : : : : : : : : : :					
Db	205 KAOMAPAEIATHY--RNAYEAKEAKEAKOQVAEAKAREIELEKSIKYEKAOTIEEKSATPA	262				
OY	183 AK----VESAKAVTTQREYV-----RFSAVERLTK--EDIKYTNKANNDKVLEKYT	229				
	: : : : : : : : : : : : : : : : : : :					
Db	263 AESASOSVEYAKGEVGAKAEALNALDPAVERLEKASEELLKDONYLRDTVNSLKEGA	322				
OY	230 LSEDRSATVELYSMLAKQTYTVNVNKYGKEVA-----VSLSEAKTIEMAQTVVADBP	285				
	: : : : : : : : : : : : : : : : : : :					
Db	333 TEEOKRAKKEEKAKIS-----EEVAKKEASAQAFKIEERANYEANKIAENHP	373				
OY	286 TALQFTVDENDETENVSPGIEFYTPAA-EKINAKGEITLAKGSTVYAAYKKKGKVVYA	344				
	: : : : : : : : : : : : : : : : : : :					
Db	374 NT--NVTTEANKAKASTIKASTEATIKASTEALINASTEATIKPSSKANYNK---KKIDEAK	428				
OY	345 ESKEEVKSAEGAAVAASISMNTVAEQNKADFTSKDEKONNNKYEGDNAYVQVELKDOPNAV	404				
	: : : : : : : : : : : : : : : : : : :					
Db	429 AAKEEKKAKTEAYIALPYTKAMAARKEKAKKSAE-----AADKAKQAQAAVANGASCK	479				
OY	405 TTGKEYESLNTLEVAVDKATGYVLVLSGK--AAVYKTVKRSKSGKALVSHIVLEIAFRQ	462				
	: : : : : : : : : : : : : : : : : : :					
Db	480 TKKDAEHATKAN-----EKKHTLETADAARKNAEKKEEEDNVAKN-----EEKMK	527				
OY	463 KAMRKIKTEKTVALSTRDVTLKVPALPDYGGREFAPRVYK--LDROCKELKEOKL	520				
	: : : : : : : : : : : : : : : : : : :					
Db	528 KKVDVY-IEKVLEAKSEE-----DTYQAOIQAEIIVQVANVEEACEKAKTAEO	575				
OY	521 EAKYVRELVTNAAQOEAGNVTNYVTJAKSGKEAKATLALIELABGAFSKEPFVRIGDTEL	580				

Db 576 EKKAKDEAV--KAKEAE-----AKKQAEKAKIT----- 605
Oy 581 DKYVEENKNAATVSVLPDANGVLKGAEAELKVTITNEKEKEDATDAQVTONNS 640
Db 606 -KTAEE-----ANKAKEEAKASEAKQAEATAGD-----VDEEVAVNE 646
Oy 641 VTTVCGAKAGETKYVTVLDDGLITTHSFKVVDPAPTKAGLAVEFTSTLKEVAPNADL 700
Db 647 FESVAAAAAHHKVPETLD-----KEKKNAENAKKASAKATEAKTTAETATKATEA 701
Oy 701 KALLNITLSDVDPATKATASNEFVSADINVAENGTGA-----RGATSIYVK 752
Db 702 KTA-----AGNAKASENKAIPAD--VLAEKASTEAKSLKEAKKLADIKRS 748
Oy 753 NLTVYKDGKQKVEFDKAVQAVSIAKAPKA 783
Db 749 NVTNEKAKRDKAANDAAHQASLSASKAKEA 779

RESULT 13
ID 086999 PRELIMINARY; PRT: 1036 AA.
AC 086999;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
DE S-LAYER PROTEIN.
GN SLPA.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIB 10682;
RX MEDLINE=98129094; PubMed=9467913;
RA Lemaire M., Miras I., Gounon P., Beguin P.;
RT "Identification of a region responsible for binding to the cell wall
RT within the S-layer protein of Clostridium thermocellum.";
RL Microbiology 144:0-0(0).
DR EMBL: U79117; AAC33404.1;
DR InterPro: IPR001119; SLH.
DR Pfam: PF00395; SLH; 1.
SQ SEQUENCE 1036 AA; 113329 MW; 1AF2202060D7C2A1 CRC64;

Query Match 6.9%; Score 268; DB 2; Length 1036;
Best Local Similarity 20.5%; Pred. No. 0.00044;
Matches 217; Conservative 133; Mismatches 325; Indels 382; Gaps 50;

Oy 16 DSINY-----LVEKGVGNDKGMPEPEKE--LTRAETATMAOI-----LNLPIDK 60
Db 27 DSEFYEKEAEILYRLGLYKGTSETSEYVPLNEGLDKQGTGMALRLFGEDDALLETPMDE 86
Oy 61 DAK---PSTADSO--GQWTPPIA-AVEKAGYIKGTGNG-FEENKIDRVSMAQLVE-- 111
Db 87 AAOITLAAKFKDAADLADMAQROVAAYAVEK-GYVKGYPDGTFLPNADLNGIAFCSLILOOL 145
Oy 112 -----AYKIDT--KVNCTPATRKDLETLN-----WGKEKA 140
Db 146 GYDGDVFDEDAAYKLOEFGGLTAEQAEAFNNKNGINRDSMVGIAFSALQAVYKATGTVI 205
Oy 141 NTLVELG-----ISVGTGQMEPKKTVTKAEAAQFIKTDKQGTFAKAVESAKAV----- 191
Db 206 EYLVENGANSKRELATLFLGLVLAKEVKAIDAVK-----QVGRPEVLPPEVEEYVEED 259
Oy 192 TTQKVEVKE-----SKAVEK--LTKEDIKVTN--RANND 221
Db 260 TTEKLAWEPTVDTSEVGEIEGTIKGASGLAYREPKATLKVITPEELQVVDVAKPUL 319
Oy 222 KVLVE-----VTLSEDKRSATYELVSLNAKQITTYVDVK 257

Db 320 KEIVIEFNGEVASKADEKSSVEDNTIELVYSEDKTTVTILVAGAMTAEEIEVTI-- 377
Oy 258 VGKTEVAVGSLKAKTIEADQGTVADEPTAL-----QFTVK-----DENGTE 299
Db 378 --KTATGLKEEVTKEFVPADENPEAESIALIGPNSFEIKFSEPYOSSDAEVLVNDGY 435
Oy 300 VVSPE-----GIEFVTPAEKINAKG-----BITLAKGS--TT 331
Db 436 YVSEELKSDYPTLVELLGVSSINDEGTYYKVGKGRDVAGNIMRTKTEFLRYKDTTPPT 495
Oy 332 VKAAYKKDKGVAAESKEVYV-----SAEGAFA-----SISNWTVAEQKAPFTSKDFQ 381
Db 496 AYKKEATQNKVYIEFNEPATRGGYSGDEALTRDFQYQYSSMKPKVYASD----- 547
Oy 382 NKKVY-----EGDNAYVVELKQDFNAVTGKVEYESLNTAEVAVDKAQKVTYLSAG 434
Db 548 NKKVTLYSESDQNDQGY-----PVYLLPVG 573
Oy 435 KAPVYTVVDSKGLAYSTVTELEAFQAKAMDKLEKTNVALSTKDYTDLYKAPVLDQ 494
Db 574 NVTITI-LKEVDDAVY-----DAGNKLSDLKLTAI--VAADNAPTVKSVTAEDK 625
Oy 495 YKEFTAPVTYK-----VLDDGKLELQKLEAKYVNRRELVNAAQAGHANTVVL 546
Db 626 IYVFESEDVNEQAKDKMYYIKKQCKEID-----TAISS 660
Oy 547 AKSGEKAATATLAEIKADGAFSKEVRGL-DTELD---KVTY---EENKNAATVSVL 598
Db 661 IYDYSNETKVTIYLDKELTSGKRYTIDIKGITSSENEKAVTIEFTEYTDAPTIEEV 720
Oy 599 PDANGVLKGAEAELK--VTTNKEG--KEYDATQAOVTY-----QNNSVITVQ 646
Db 721 TFVDNTIYVYSEAMSTKNGSVLKNQKYLVDNDKVEIKKIELEFGSDKMKVRLTVS 780
Oy 647 GAKAGETVYVY-----VLDDKLTTHSF-----KVV 673
Db 781 DVDLNDVDELTITANVDEAGNNAISAFDVAKKLSQAEVSEIRISTHEIYINKLL 840
Oy 674 DPAAPAK-----GLAVEFTSTLKEVAPNA-----DLKALLNLT 708
Db 841 DKATVEKTFEVEGRGSKNVALTRISSITVDGKTIYKGLPDAVRPANGDITGYLYI- 899
Oy 709 SVDGVAATTA--ATASNVEFVSADINVAENGTG---AKGATSIYVKNL----- 754
Db 900 -VGEIKSDTKEMAGAVAKPVDKFPSPFSVANGVYGDASKGFTLTFEDIKFLNNS 958
Oy 755 -----TYVKDGK---EOKVEFDKAV---QVAVSIX 778
Db 959 AGLGATDLYIKKSKTLEAGIDYDVAIDNKITVTILK 995

RESULT 14
ID 053505 PRELIMINARY; PRT: 1109 AA.
AC 053505;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DT 01-NOV-1998 (Tremblrel. 08, last annotation update)
DE TYPE A S-LAYER PROTEIN.
GN SLP2.
OS Campylobacter fetus.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=196;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=95204338; PubMed=7896695;
RX Dworkin J., Tummuru M.K., Blaser M.J.;
RT "A lipopolysaccharide-binding domain of the Campylobacter fetus S-
RT layer protein resides within the conserved N terminus of a family of
RT silent and divergent homologs.";
RL J. Bacteriol. 177:1734-1741(1995).
DR EMBL: 576860; AAB33871.1; -.


```

Db 1274 -----TVAESTEPSSTF-----AITSUK----- 1292
QY 416 TEVAVVDKATGKVTVLKAGKAPKVTVKDSKGLVSHTEIEAPAKAMDKIKLEKTNV 475
Db 1293 TKMSTISEETGKTSVSE-----EITVKT-----TWTEESTEPSSTEAITSEKTSV 1337
QY 476 ALSTKDVV-----DLVKAPVLDQYKGEFTAPYTVKVLDDKDKELKQKLEAKYVNR 528
Db 1338 -ISTSEETGKTSVSEEVTKTVTDE-ATEITSTVSI-----ETKE----- 1376
QY 529 LVLNAGOEAGNYTVLTAKSKEKAKATLLELKAPGAFSKFEVRLDTELDKVTVEEN 588
Db 1377 -----TSVSGSTELST-----QASSKIESPTESGITSHTHESE 1411
QY 589 QKNMPTVSVLPVDANGLVL-KGAFAELKVTTNKEGKEVDATDAQVTVQNNSVITVGOG 647
Db 1412 ESTVSTTEKGEVTSSETTELKTSIVSTEMLSTTEKE----- 1447
QY 648 AKAGETVKTVVLDDKLTITHSFKVVDTAFTAKGLAVEFTSLKEVAPNADLKALLNI 707
Db 1448 -----ETVWSTTELSTEEF-----SSTESTETGSTE----- 1477
QY 708 LSVDGVPATTAFAKATFASNTEFVSADTNVVAENGTGAKGATSIYKNLTVVKDKGEQKVEF 767
Db 1478 ISTETESTTAS-----QIIPASTTVAPEINETGTP-ASALPVEGLSTSTEG-----ITT 1526
QY 768 DKAVOVAVSIREAKPAT 784
Db 1527 QESEVTVITTEKSKRGT 1543

```

Search completed: March 23, 2002, 08:01:55
 Job time: 628 sec

378 TTCTCAAGTAAACCTTGGGCTTAAATAATCTTAGAAATCGCAGTAGTGA 427
146 eugliiIleSerValGlyThr... GlyAspGlnTrpGluProIlysthr 161
428 ACGGCAATTTTCGAAGTACTGATGCAACAACTTAAACCTAACCACTCA 477
162 ValThrLysAlaGluAlaGlnPheIleAlaLysThr... AspLysGln 177
478 ATCACTCGTCAGACACTTTCGACTAGTGTTCAACAGCTACAGTGTACAAAGT 527
177 nPheGlyThrGluAlaAlaLysValGluSerAlaLysAlaValThrThrG 194
528 TGAAGCTGAACACTCCAGAAGAAGCAGCATTTGTTAAAGCTATCAACACA 577
194 LuysValGluValLysPheSerLysAlaValGluLysLeuThrLys... 209
578 CAACCTGTGAAGTACACTTTCGAAGAAGAGTACTTACGTTCAAGCATT 627
210 GluAspIleLysValThrAsn..... LysAlaAs 219
628 AACTTCAAAATCGAAGCTTTTGAATAATTAATTCGTTGTTAAACAAC 677
219 nasAspLysValLeuVal..... LysG 227
678 AAACAAAAAGTTGTTCTATTACTTAACTGACAGCTCAACAGCTGATTAAG 727
227 LuValThrLeuSerGluAspLysArgSerAlaThrValGluLeuThrSer 243
728 AGATGCTTTTAACCTTTCGAC..... GCGGAACACATCGGTGGCTTTAA 771
244 AsnLeuAlaAlaLysGlnThrThrValAspValAsnLysValGlyLys 260
772 GGGTGGCTGCT..... GTAGTCCAACTAAAGTTGGAAC 806
260 sThrGluValAlaVal... GlySerLeuGluAlaLysThrIleGluMetA 276
807 AGATCTTTCGCGAGTTCGAAGGTCAATCTTAAGCTGATTCCTGTTACT 856
276 IaAspGlnThrValValAlaAspGluProThrAla..... LeuGln 289
857 CTAAAGTAACTGTTGCTGAAGGTCAATCTTAAGCTGATTCCTGTTACT 906
290 PheThrValLysAspGluAsnGlyThrGluValLysSer..... 302
907 TTCAGTGTACAGTAAACAATGATGGCTGTATACCAACATTTAACAG 956
302 302
957 TGAAGCTTTAAACAAGCAAGAGGTATCGCAACATCTCTACACCTCT 1006
303 .. ProGluGlyIleGluPheValThrProAlaAla..... 313
1007 ATTAAGCAAGTACTGATGAAGTAACTGCTTATGCAACTGGTGATGCTTCT 1056
314 GluLysIleAsnAlaLys 319
1057 AAATTTCTCACTGGTTATGTATTTCTGGGTAGATACCAATTCCTTCACT 1106
319 sGlyGluIleThr..... LeuAlaLysGlyThrSerThrThrV 332
1107 TGAAGAGTAACTACAGTGTCTGATTAATGATGGTCAACAAACACTT 1156
332 alLysAlaValThrLys..... LysAspGlyLysVal..... 342
1157 ACAAAAGTACTTATTAATAAACCTTAATAACTGTAAACCGAGAACCAACAA 1206
343 ValAlaGluSerLysGln 348
1207 ACATTTAAATGTTGGTTTGTAGAAAAACATGATGTTCATTCTTGATTAAGT 1256
348 uValLysValSerAlaGluGlyAlaAlaValAlaSerIleSerAsnTrpT 365
1257 ACGAAATGCTACAGTTATGGCGTTAAAGACATTACAAATTAAGCAATGTA 1306
365 hrValAlaGluGlnAsnLysAlaAspPheThrSerLysAspPheLysGln 381
1307 CAGCTTTAGAC..... GCTGCTCAAAATTAACACAGATTTCTTAAGGTGAA 1350
382 AsnAsnLysValIleGlyGluLysAspAsnAlaThrValGluValGluLeuLys 398
1351 GCTACATTACAGATTTCGTGACTACTAATGACAGCTGTAACGCCAGTAGTAA 1400
398 sAsp..... GlnPheAsnAla. 403
1401 TGATCTACACAGCAGCTAACAATAGTACTTCAAAATTAATAATATAGTCAT 1450
404 ValThrThrGlyLysValGluIleGlySerLeuAsnThr 416
1451 CTGCTTTCAAAACTGCTTCTTAAGTAAAGTACTTTCGCTGCTTCAAGCA 1500
417 GluValAlaVal..... ValAspLysAlaThrGlyLysValThr 429
1501 GAGTATACAAATTTGAGTTACCTGCGTGTATATGCT... GGAGAAGTTGC 1547
429 rValLeuSerAlaGlyLysAla... ProValLysValThrValLysAsp 445
1548 TGCAATTTGGCCTACTAAGCGTCCGCAATACAAAGTTATGTGAAGAATA 1597
445 eLysGlyLysAlaLeuValSerHisThrValGluIleGlu..... 458
1598 AAGCTGTAACTTACGCTAAATGAATGAACGTTAATGTTGCACTTCAATGAA 1647
458 458
1648 GATAAAGTCGTGAATTTCAACAGCTTACAAATGCTAAATTCGTGTATAC 1697
459 AlaphaIleGlnLysAlaMetLysAspIleL 469
1698 TGATCCAGATCTGCGATATCTTACACAGCGCATTAAGCAAAACAATCT 1747
469 yLysGluLysThrAsn..... 474
1748 CTGTA... AAAACAATGATTAAGGTGAAGCTACATTTGTATGCTTCT 1794
475 ValAlaLeuSerThrLysAspValThrAs 484
1795 GATACAGTAACGATATGCAACACCAATTCCTGGATGTATATATATAC 1844
484 pLeuLysValLysAlaProValLeuAspGlnThrGlyLysGluPheThrA 501
1845 TTCTGATCAAAACAAAGCGACGCTTGAATGAAGGTGAACCAAAACAGATG 1894
501 IaProValThr... ValLysValLeuAspLysAspGlyLysGluLeuLys 516
1895 CACCAATCTCTTACTTCCAAACACCATATCTTGTAGGCTGCGTATGAAA 1944
517 GluGlnLysLeuGluAlaLysThrValAsnArgLysLeuValLeuAsnAl 533
1945 GCATACAAA..... AAATCAGATCTTAAATAAGCTGAACATAAGTT 1985
533 aAlaGlyGlnGluAlaGlyAsnThrThrValValLeuThrAlaLysSerG 550
1986 TGATGCTTCAAACTGACAGATTTTGCAGCAATTTAGTAACCAAGCG 2035
550 LysGluLysGluAlaLysAlaThrLeuAlaLeuGluLeuLysAlaProGly 566
2036 GCAAAAAGTACGTTACTTCTTAT..... AAGAAAGCA 2070
567 AlaPheSerLysPheGluValArgGluLeuAspThrGluLeuAspLysTy 583
2071 ACTTATCAATCTCAATCTACGTCGCTCAATGATATTTAAAGTACATACCA 2120
583 rValThrGluGluAsnGluLysAsnAlaMetThrValSerValLeuProV 600
2121 AGTATATCCACCAATCTGATGCTACAGATTAATTAAGT... ACTT 2167


```

162 valThrLysAlaGluAlaLeuIleuPheIleAlaLysThr...AspLysGlu 177
   ::::: ::::: ::::: ::::: :::::
478 ATCACTCGCAAGACTTTCAGTACTGATCAAAAGCTAGTGAACAAGT 527
   ::::: ::::: ::::: ::::: :::::
177 nPheGlyThrGluAlaLeuLysValGluSerAlaLysAlaValThrThg 194
   ::::: ::::: ::::: ::::: :::::
528 TGAAGGTCAAACTCCGAAAGAACGACATTGTTAAAGCTATCAACACA 577
   ::::: ::::: ::::: ::::: :::::
194 LuLysValGluValLysPheSerLysAlaValGluLysLeuThrLys... 209
   ::::: ::::: ::::: ::::: :::::
578 CAAGCTGTGAAGTACATTGCAAGAAAGAGTACTACGCTTCACAGACT 627
   ::::: ::::: ::::: ::::: :::::
210 .....GluAspIleLysValThrAsn.....LysAlaAs 219
   ::::: ::::: ::::: ::::: :::::
628 AACTTCAAAATCGAAGTTAGAAATTAATAATGCTTGTAAACAAAC 677
   ::::: ::::: ::::: ::::: :::::
219 nasnAspLysValLeuVal.....LysG 227
   ::::: ::::: ::::: ::::: :::::
678 AAACAAAAAGTTGTTGTTAATTAAGTCAAGCTCAACAGCTGATTAAG 727
   ::::: ::::: ::::: ::::: :::::
227 LuValThrLeuSerGluAspLysArgSerAlaThrValGluLeuThrSer 243
   ::::: ::::: ::::: ::::: :::::
728 AGTATGTTTAACTCTTGAC.....GGCGAAACATCGGTGGCTTAA 771
   ::::: ::::: ::::: ::::: :::::
244 AsnLeuAlaLysGluThrThrValAspValAsnLysValGly 260
   ::::: ::::: ::::: ::::: :::::
772 GGTGGCTGCT.....GTAGTCCAACTAAAGTGAAGTCACT 806
   ::::: ::::: ::::: ::::: :::::
260 sThrGluValAlaVal...GlySerLeuGluAlaLysThrIleGluMet 276
   ::::: ::::: ::::: ::::: :::::
807 ACTATCTTCTGCAGTTCAGAGTAACTGTGTCAAGAAAGTAAAGTTCAAG 856
   ::::: ::::: ::::: ::::: :::::
276 LaAspGluThrValValAlaAspGluProThrAla.....LeuGln 289
   ::::: ::::: ::::: ::::: :::::
857 CTAAAGTACTGTGCTGAAGCTCAATCAAGCTGTATCTCTGTACT 906
   ::::: ::::: ::::: ::::: :::::
290 PheThrValLysAspGluAsnGlyThrGluValLysSer..... 302
   ::::: ::::: ::::: ::::: :::::
907 TTCAGTGTACAGTACAAACAATGATGGCTGTACCAACATTTAACAG 956
   ::::: ::::: ::::: ::::: :::::
302 .....GluLysIleAsnAlaLys 319
   ::::: ::::: ::::: ::::: :::::
957 TGAAGCTTTAACAAAGCAAGAGGTATGCCAACATACCTTACACTCGTT 1006
   ::::: ::::: ::::: ::::: :::::
303 ..ProGluGlyIleGluPheValThrProAlaAla..... 313
   ::::: ::::: ::::: ::::: :::::
1007 ATAAGAAGTACTGATGAAGTAACTGTTATGCAAACTGTGATCGTCT 1056
   ::::: ::::: ::::: ::::: :::::
314 .....GluLysIleAsnAlaLys 319
   ::::: ::::: ::::: ::::: :::::
1057 AATTCTCAGTGGTTATGTATTCGGGGGTGATACAAATTCTTTCAGT 1106
   ::::: ::::: ::::: ::::: :::::
319 sGlyGluIleThr.....LeuAlaLysGlyThrSerThrThrThv 332
   ::::: ::::: ::::: ::::: :::::
1107 TCAAGACAGTACTACAGGTGCTTCAGTTAATATGTGCAAAACAAACTT 1156
   ::::: ::::: ::::: ::::: :::::
332 alLysAlaValThrLys.....LysAspGlyLysVal..... 342
   ::::: ::::: ::::: ::::: :::::
1157 ACAAAAGTACTATATAAAACCTAAACCTGTAACCAAGCAACAAACAA 1206
   ::::: ::::: ::::: ::::: :::::
343 .....ValAlaGluSerLysGlu 348
   ::::: ::::: ::::: ::::: :::::
1207 ACATTTAATGTTGTTTGTAGAAAACATGANTGTACTTCTGATAAAGT 1256
   ::::: ::::: ::::: ::::: :::::
348 uValLysValSerAlaGluGluValAlaValAlaSerIleSerAsnTrpT 365
   ::::: ::::: ::::: ::::: :::::
1257 ACCAATGCTACAGTTAATGGCGTAAAGACATTACAATTAAGCAATGGTA 1306
   ::::: ::::: ::::: ::::: :::::
365 hValAlaGluGlnAsnLysAlaAspPheThrSerLysAspPheLysGln 381
   ::::: ::::: ::::: ::::: :::::
1307 CAGCTTTAGAC.....GCTGCTCAAAATTACAAACAGATTCTAAAGTGAA 1350
   ::::: ::::: ::::: ::::: :::::
382 AsnAsnLysValThrGluGluLysAsnAlaThrValGluValGluLeuLys 398
   ::::: ::::: ::::: ::::: :::::

```

```

1351 GCTACATTCACAGTTTCTGCTACTATGACGCTGAACCCAGTAGTATA 1400
   ::::: ::::: ::::: ::::: :::::
398 sAsp.....GlnPheAsnAla. 403
   ::::: ::::: ::::: ::::: :::::
1401 TGATCTACACAGCACTAACATAGTACTTCAAATATAAATAATTAATGCT 1450
   ::::: ::::: ::::: ::::: :::::
404 .....ValThrThrGlyLysValGluThrGluSerLeuAsnThr 416
   ::::: ::::: ::::: ::::: :::::
1451 CTGCTTTACAACTACTGCTTCTTAAGTAACTTGGCTGCTTTCACAGA 1500
   ::::: ::::: ::::: ::::: :::::
417 GluValAlaVal.....ValAspLysAlaThrGlyLysValThr 429
   ::::: ::::: ::::: ::::: :::::
1501 GAGTATACAAATGAGTAACTGCTGCTGATATATGCT...GGAGAAGTGC 1547
   ::::: ::::: ::::: ::::: :::::
429 rValLeuSerAlaGlyLysAla...ProValLysValThrValLysAspS 445
   ::::: ::::: ::::: ::::: :::::
1548 TGCAAATGGCGCTACTACGGTGGCAATACAAAGTTATTTGTAAGATA 1597
   ::::: ::::: ::::: ::::: :::::
445 eLysGlyLysAlaLeuValSerHisThrValGluIleGlu..... 458
   ::::: ::::: ::::: ::::: :::::
1598 AAGCTGTAACTTACTTAATAATGAATCGTTAATGTCATTCATGA 1647
   ::::: ::::: ::::: ::::: :::::
458 ..... 458
   ::::: ::::: ::::: ::::: :::::
1648 GATAAAGATCGTATATTCAACAGTTACAAATGCTAAATTCGTGATAC 1697
   ::::: ::::: ::::: ::::: :::::
459 .....AlaPheAlaGluLysAlaMetLysAspIleL 469
   ::::: ::::: ::::: ::::: :::::
1698 TGATCCAGTACTGCAATATCTTCAACAGCGATTAACCAAAACAATCT 1747
   ::::: ::::: ::::: ::::: :::::
469 yLysLeuGluLysThrAsn..... 474
   ::::: ::::: ::::: ::::: :::::
1748 CTGTA...AAAACAATGATAAAGTGAAGCTACATTTGTATCGTCT 1794
   ::::: ::::: ::::: ::::: :::::
475 .....ValAlaLeuSerThrLysAspValThrAs 484
   ::::: ::::: ::::: ::::: :::::
1795 GATACAGTAAACGATTATGCAACACCAATGCTGATGATTATTAATAC 1844
   ::::: ::::: ::::: ::::: :::::
484 PheLysValLysAlaProValLeuAspGluThrGlyLysGluPheThr 501
   ::::: ::::: ::::: ::::: :::::
1845 TTCGATGCAAAACAAGCGACCTGTATGAAGTGAACCAAAAGCAGTGG 1894
   ::::: ::::: ::::: ::::: :::::
501 LaProValThr...ValLysValLeuAspLysAspGlyLysGluLeuLys 516
   ::::: ::::: ::::: ::::: :::::
1895 CACCAATCTCTTACTTCCAGACCATATCTTGAAGCTGACGTACAA 1944
   ::::: ::::: ::::: ::::: :::::
517 GluGlnLysLeuGluAlaLysThrValAsnArgGluLeuValLeuAsnAl 533
   ::::: ::::: ::::: ::::: :::::
1945 GCATACAAA.....AAATCAGATCTTATATAAGCTGTAACTTAAGT 1985
   ::::: ::::: ::::: ::::: :::::
533 AlaGlyGlnGluAlaGlyAsnThrThrValValLeuThrAlaLysSerG 550
   ::::: ::::: ::::: ::::: :::::
1986 TGATGTTCTGAACCTGCAATTTTTCAGACGAAATTAAGTAAACCAAGC 2035
   ::::: ::::: ::::: ::::: :::::
550 LyGluLysGluAlaLysAlaThrLeuAlaLeuGluLeuLysAlaProGly 566
   ::::: ::::: ::::: ::::: :::::
2036 GCAAAAAGTAACTGCTACTCTATTT.....AAGAAGCA 2070
   ::::: ::::: ::::: ::::: :::::
567 AlaPheSerLysPheGluValArgGlyLeuAspThrGluLeuAspLysTy 583
   ::::: ::::: ::::: ::::: :::::
2071 ACTTATACAAATCTCAAAATCGGTGCTAATGATATTAAGTAGATMAACA 2120
   ::::: ::::: ::::: ::::: :::::
583 rValThrGluGluAsnGlnLysAsnAlaMetThrValSerValLeuProV 600
   ::::: ::::: ::::: ::::: :::::
2121 AGTTATCTCACCAAATGCTAGCTACACAGTACTATATAGACT...ACTT 2167
   ::::: ::::: ::::: ::::: :::::
600 AlaSpAlaAsnGlyLeuValLeuLysGlyAlaGluAlaValGlu..... 614
   ::::: ::::: ::::: ::::: :::::
2168 TATCTTTCACAGAACTGTATTAACCTGCTAAGAAATTAAGAGTTACT 2217
   ::::: ::::: ::::: ::::: :::::
615 .....LeuLysValThrThrThr 620
   ::::: ::::: ::::: ::::: :::::

```



```

637 .....GlnAsnAsnSerValIleth 643
2552 ACATCAATCAATACAGCTGATTCGTGTTCAACGCAACTCTATTGGTT 2601
643 rValGlyGln.....GlyAlaLysAlaGlyGluThrTyrL 655
2602 TCGTGGTAAACCAAGTGAATATGCTGGTGTATCGACGCAAAACATATA 2651
655 ys..... 655
2652 AATACCTCGAGCTAATGTAATGAGTATTGGTGAAGCGCATGGAA 2701
656 .....ValThrVa 658
2702 GCATTATTAACTCAATATGCAACTGAAGCCAAAAGTAAACAATCTCATA 2751
658 lValLeuAspGlyLysLeuIleThrHisSerPheLysValValAspT 675
2752 TAATGTGATGGTGTATACAGTTACA.....TTTAAAGTAAATTAGT. 2791
675 hrAlaProThrAlaLysGlyLeuAlaValGluPheThrSerThrLeu 691
2792 .....GCTGTATATCTTCACACTGAAGCTATC 2818
692 LysGluValAlaProAsn.....AlaAspLeuLysAlaAlaLeuLe 705
2819 AAACCACTTCTCCACACACGACGCTCCACTACTGCGCATTTAC 2868
705 uAsnIleLeuSerValAspGlyValProAlaThrThrAlaLysAlaThra 722
2869 ATTAACACCGACAGCTGTGTTAGTTAGTTTAACTGACATTAACA 2918
722 lAserAsnValGluPheValSerAlaAspThrAsnValValAla..... 736
2919 CTTAGGAATTTCTCATTTAGCTGATGACATCTTAATGTAAGCAACAAC 2968
737 ..GluAsnGlyThrValGlyAlaLysGlyValLeuThrSerIleTyrVal 752
2969 GTTGATACCTCAACTGTTCTCAATAAAGATGCAATATTCATTA.. 3016
752 sAsnLeuThrValValLysAspGlyLysGluGlnLysValGluPheAspL 769
3017 .TCTCTACATTAAGTTGAACGTGCTAATACAGTGTA...TTTGCTA 3062
769 yAlaValGln.....ValAla 774
3063 CAACGTGTTCAAGCTGTACATTATCTTTAACTGCTGTACATTAAACA 3112
3113 GTTACTTATGCAGATGCTAAATAATGCT 3139

seq_name: /c9n2_6/plodata/2/lna/5B.COMB.seq: US-08-682-517-8

seq_documentation block:
; Sequence 8, Application US/08682517
; Patent No. 5874267
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/682,517
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4197 base pairs

```

```

;
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Bacillus sphaericus
; INDIVIDUAL ISOLATE: P-1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 95..3850
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 185..3850
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 95..184
; US-08-682-517-8

```

```

alignment_scores:
  Quality: 378.50      Length: 1009
  Ratio: 0.762        Gaps: 44
  Percent Similarity: 49.257  Percent Identity: 22.101

```

alignment_block:

US-09-754-947-1 x US-08-682-517-8 ..

Align seg 1/1 to: US-08-682-517-8 from: 1 to: 4197

```

16 AspSerIleAsnTyrLeuValGluLysGlyAlaValLysGlyAsnAspLys 32
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
227 GAAGCACTTCAAGCTTATGTTGACCAAGCGCTAATCCAAAGTGTACTTAA 276
32 sGlyMePheGluProGlyLysGluLeuThrArgAlaGluAlaLeuThrM 49
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
277 CGGGAACCTTCAACCCACTTAACACAGTAACACTGTCGCAACAGCTGCAAAA 326
49 etheAlaGlnIleLeuAsnLeuProIleAspLysAspAlaLysProser 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
327 TCTTCACAAAAGCTTGAATTAAGACCTTAACGTAACGAGATGTA.....AAC 370
66 PheAlaAsp...SerGlnGlyIntRPTyrThrProPheIleAlaAlaVa 81
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
371 TTCAAAGACGTGAAGCTGGCGCTGTGTACTACACTCAATCGTGTGCTGT 420
81 lGluLysAlaGlyValIleLysGly...ThrGlyAsnGlyPheGluProA 97
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
421 TGTAGCTAACGCGCATTTTGAAGGTGTAGTGCACAACTGAATTCACACCA 470
97 snGlyLysIleAspArgValSerMetLaserLeuValGluAlaTyr 113
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
471 ACAAAATCTTAACCTGTTCTGAAGCTGCTAAAAATTTAGTAGAAGCATTC 520
114 LysLeuAspThrLysValAsnGlyThrProAlaThrLysPheLysAspLe 130
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
521 GGTTTAGAGGTGAAGCAGAT.....CTTAGCCAAATTTGCTGACCC 561
130 uGluThrLeuAsn...TrpGlyLysGluLysAlaAsnIleLeuValGluL 146
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
562 TTTCACAGTAAACCTTGGGCTAAATAATCTAGAAATGCGACTACTCTA 611
146 euGlyIleSerValGlyThr...GlyAspGlnTrpGluProLysLysThr 161
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
612 ACGGCATTTTCGAAGGTACTGATGCAAAACAACTTAACCTTACCAACTCA 661
162 ValThrLysAlaGluAlaAlaGlnPheIleAlaLysThr...AspLysG 177
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
662 ATCACTGCTCAAGCTTTCAGTACGTGTTCAAAAGTACAGTTGCAACAGT 711
177 nPheGlyThrGluAlaAlaLysValGluSerAlaLysAlaValThrThrg 194
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
712 TGAAGGTGAAGATCCAGAAAGACAGACAGATTTGTTAAAGCTATACACACA 761

```

194 IntyValIGlValILysPheSerLysAlaValGluLysLeuThrLys... 209
195 ||||| |||
762 CAACGTGTTGAAGTAAACATCTGGAAGAAAGAACTACTAACCTTCAACACTT 811
210 GlusPILeLysValIThrasn LysAlaAs 219
812 AACCTCAAAATTCGAAGCTTTAGAAATTAATAAAATGCTTCTGTTAAACAAC 861
219 naaNaAPLysValLeuVal LysG 227
862 AAACCAAAAAGGTGTGTTACTTAACTAGCAAGCTCAACAGCTCATGAATAG 911
227 LuValIThrLeuSerGluAspLysArgSerAlaIThrValGluLeuTyrSer 243
912 AGATATGTTTTAACTCTTGAC GCGAACAATGCGTGGCTTTAAA 955
244 AsnLeuAlaLALysGlnIThrTyrIThrValAspValAsnLysValGlyLys 260
956 GGTGTGGTGGCT GAGTTCAACATGAAGTGAAC 990
260 sThrIGluValAlaVal... GlySerLeuGluAlaLysThrIleGluMetA 276
991 AGATATCTCTGCGATTCAAGGAATACCTGGTCAAAAGATGAAGTTCAAG 1040
276 IaaspGlnIThrValValAlaAspGluProCthAla LeuGln 289
1041 CTAAAGTAACTGTTGGCTGAAGTCAATCAATCAAGCTGATTCCTGTACT 10990
290 PheIThrValLysAspGluAsnGlyIThrGluValLysSer 302
1091 TTCAGTATCCAGGTAAACAATGATGGCGTTGTACCAACATTAAACAG 1140
302 302
1141 TGAAGCTTAAACAAGCAAGAGGGATATGGCAACATCTTACACTGTT 1190
303 .. ProGluGlyIleGluPheValIThrProAlaAla 313
1191 ATTAAGAAAGCTACTGATGAAGAACTGCTATGCAACGTGATGCTGTTCT 1240
314 GlusITLeuAlaLys 319
1241 AATATCTCACTGGTATGATATTCGGGGTGAAGTATCAATCTCTTCACT 1290
319 sGlyGluIleThr LeuAlaLysGlyIThrSerIThrV 332
1291 TGAAGAAAGTAACTACAGAGCTTCAAGTTAAATTAAGTGCACAAACAACTT 1340
332 aLysAlaValLysLys LysAspGlyLysVal 342
1341 ACAAAAGTACTATTAACAAACCTTAAGCTGAACCAAGAACAAACAA 1390
343 ValAlaGluSerLysG 348
1391 ACATTTAATGTTGGTTTGTGAACAAACATGAATGATTTACTTGTATAAAGT 1440
348 uValLysValSerAlaGluGlyAlaAlaValAlaSerIleSerAsnITrPT 365
1441 ACCAAATGCTACGTTAATATGGGTAAAGCACTTCAATTAAGCAATGCTGA 1490
365 hValAlaGluGlnAsnLysAlaAspPheIThrSerLysAspPheLysGln 381
1491 CAGCTTTAGAC GCTGCTCAAAITTCACACAGATTTCAAAAGTGCAA 1544
382 AsnAsnLysValITyrGluGlyAspAsnAlaITyrValGlnValGluLeuLys 398
1535 GCTATACCTACAGATTTCTGGTACTAAATGCAAGCTTAAAGCCAGTACTATA 1584
398 sAsp GluPheAsnAla 403
1585 TGATCTACACAGCACTAAACAATAGTACTTCAAATTAATAAAATATATGTCAT 1634
404 ValIThrIThrGlyLysValGluITyrGluSerLeuAsnIThr 416

[illegible]

```
2502 CTACATTCACAGCTACAAATGAAGTCCAACTCTTACACTGTGTAGCT 2551
637 .....GlnAsnAsnSerValIleTh 643
2552 ACTCAATTCATACAGCTGATTCTGTTCACAAACAGCAACTCTAATTGGTT 2601
643 rValGlyGln.....GlyAlaLysAlaGlyIleThrTYR 655
2602 TGCTGTGTAACAAACCCAGTAATAATGCTGTGTATCAGCCAAACATATA 2651
655 YS..... 655
2652 AATACCTCGAGCTAATGTATGATGAAGTATTGGTGAAGCGCATGGCAA 2701
656 .....ValThrVa 658
2702 GCATTATTAACTCAATATGCAACTGAAGCCAAAGTAACAACTCTCTA 2751
658 lValLeuAspGlyLysLeuIleThrHisSerPheLysValAlaAspT 675
2752 TAAATGTAGATGTGATACAGTTACA.....TTTAAAGTAAATTACT. 2791
675 hValProThAlaLysGlyLeuAlaGluPheThrSerThSerLeu 691
2792 .....GCTGTTAATTCTTCACTGAAGCTATTC 2818
692 LysGluValAlaProAsn.....AlaAspLeuLysAlaAlaLeuLe 705
2819 AAACCAAGTTGCTCAACACAGCAAGCACTCCCACTACTGCGCATTTAAC 2868
705 uAsnIleLeuSerValAspGlyValProAlaThrThAlaLysAlaThra 722
2869 APTAACACCAACAGCTGTGTGTTAGTTAACAACACTGACACTAACA 2918
722 lAsnAsnValGluPheValSerAlaAspThrAsnValAla..... 736
2919 CTTTAGCAATTCATTAAGTCAGATCTTAAATGTAAGCCAAACACT 2968
737 ..GluAsnGlyThrValGlyAlaLysGlyAlaThrSerIleThrVally 752
2969 GTTGATACTGCAACTGTTTCATTTAAAGATAGTGCAAAATTAATTCATTA. 3016
752 sAsnLeuThrValValLysAspGlyLysGluGlnLysValAluPheAspL 769
3017 .TCTCTACATTAAGTTGAACCTGGTCAATATACAGGTGTA...TTTGCTA 3062
769 yAlaValGln.....ValAla 774
3063 CAACCTGTTCAAGCTGTACATTAATCTTAACTGCTGTACATTACA 3112
775 ValSerIleLysGluAlaLysProAla 783
3113 GTTACTTATGCAAGTCTTAAATAATGCT 3139

seq_name: /cgn2.6/prodata/2/1na/5A_COMB.seq:US-08-119-125A-2

seq_documentation_block:
; Sequence 2, Application US/08119125A
; Patent No. 5610011
;
; GENERAL INFORMATION:
; APPLICANT: SMITH, Hilda Elizabeth
; TITLE OF INVENTION: DNA Sequences which code for Virulence
; TITLE OF INVENTION: Characteristics of Streptococcus suis and parts thereof, poly
; TITLE OF INVENTION: antibodies derived therefrom and the use thereof for the diag
; TITLE OF INVENTION: protection against infection by S. suis in mammals, including
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Centraal Diegeneeskundig Instituut
; STREET: Edelherweg 15
; CITY: PH Lelystad
; STATE: The Netherlands
; COUNTRY: The Netherlands
```

```
; ZIP: NL-8219
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS v.6.0
; SOFTWARE: Wordperfect v. 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/119,125A
; FILING DATE: 20-SEP-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/NL92/00054
; FILING DATE: 19-MAR-1992
; APPLICATION NUMBER: NL 9100510
; FILING DATE: 21-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Handal, Anthony H.
; REGISTRATION NUMBER: 26275
; REFERENCE/DOCKET NUMBER: SMITHHE119125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (203) 838-8589
; TELEFAX: (203) 838-8794
; INFORMATION FOR SEQ. ID NO. 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6744 base pairs
; TYPE: Nucleic acid with corresponding amino acids
; STRANDEDNESS: single stranded
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus suis type II (pathogenic)
;
; FEATURE:
; OTHER INFORMATION: Extracellular factor related protein (EF*) gene
;
; FEATURE:
; NAME/KEY: promoter -35 region
; LOCATION: bp 66 to 71
;
; FEATURE:
; NAME/KEY: promoter -10 region
; LOCATION: bp 89 to 94
;
; FEATURE:
; NAME/KEY: promoter -35 region
; LOCATION: bp 153 to 158
;
; FEATURE:
; NAME/KEY: promoter -10 region
; LOCATION: bp 176 to 181
;
; FEATURE:
; NAME/KEY: ribosome binding site
; LOCATION: bp 350 to 356
;
; FEATURE:
; NAME/KEY: signal peptide
; LOCATION: bp 361 to 498
;
; FEATURE:
; NAME/KEY: start of repetitive units R1-R11
; LOCATION: bp 2869, 3097, 3292, 3520, 4087, 4381, 4609, 4837,
; LOCATION: 5065, 5293, 5521:
;
; FEATURE:
; NAME/KEY: start of repetitive Asn-Pro-Asn-Leu sequences
; LOCATION: bp 2932, 3160, 3355, 3583, 4150, 4444, 4672, 4900,
; LOCATION: 5128, 5356, 5584:
;
; FEATURE:
; NAME/KEY: dyad symmetry regions
; LOCATION: from bp 6554 to 6566 and from bp 6571 to 6583
;
; FEATURE:
; NAME/KEY: dyad symmetry regions
; LOCATION: from bp 6611 to 6625 and from bp 6631 to 6644
;
; US-08-119-125A-2
```

```
alignment_scores:
Quality: 245.50 Length: 947
Ratio: 0.557 Gaps: 41
Percent Similarity: 46.568 Percent Identity: 21.225
```

alignment_block:
US-09-754-947-1 x US-08-119-125A-2

Align seg 1/1 to: US-08-119-125A-2 from: 1 to: 6744

15	IIleaspeSIIleasEnrryIleuValIGluValysGIyAlaValysGIyAsnAs	31
3100	CTTATGCTCTTTGAATTAAGCTT...GAAAGATATAGCGAAGAA...AC	31400
31	plysGIyMetPheGluProGlyLysGluLeuThrArgAlaGluAlaIAT	48
3141	TAAAGCGACGATTAATGATGCTTAATCCAAACTTAACCTCGGAAGAGAAAGGA	31900
48	hrMetCLeIaGlnIleLeuAsnLeu...:::56	
3191	AAGCTATTGCTTAAGCTTAAGTAAGACCTGTTAAATGCTGAATCGACATT	32400
57ProIIleasplysaAspAlaIalysProSerPheAlaAs	68
3241	TTGTGCAAGCCTACCCACAGAAACAGTTCAGTACGAGGAGATTAAGCTCGA	32900
68	pserGIyGlyGlnTrpTyrThrProhelleAlaIalValGIyAlaG	85
3291	CAAAAGAT.....CTTGCCAAATTAACACTTCAG	3319
85	IyAlIleIysGlyThrGlyAsnGlyPheGluProAsnGIyIle...100	
3320	CAGCAGCAGACGCGTGGCAGAGAAAGCATTGAAGCAATTCGAATTTGACT	3369
101AspArgValSerMet	105
3370	CCAGAAAGCAAAAGATGTACCTAAGCAAGCAGTAGAAGACGCGGTTAAAGT	3419
105	AlaSerLeuLeuValGluAlaIalTyrIlysLeuAspThrIlysValAsnGIyT	122
3420	GCGGACAGACGCTATTGATTAAGCGCTCAACTCCAAACCGAATGTACACAG	3469
122	hrProAlaThrIlysrPheIysAspLeuGluThrIleAsnTrp...135	
3470	CGACAAAGCATGGATGAAGAGCTATTGATTCGACAGAAAGCTTAAAGCTACT	3519
136	GlyIysGlyIysAlaAsnIleLeuValGluLeuGlyIleSerValGly..151	
3520	CAGAAAGATGCTTAAGCAAGATTTGCCAAAGAACCAAACTACAGCTTAAGAA	3569
152ThgIyAspGIyTrpGluProLysL	160
3570	AGCGATTCAGCGCATCCAAACTTGACTCCAGATGAGAAAGCATTCAGCTA	3619
160	ystrValThrIlysAlaGluAlaIalGlnPheIleAlaIalysThrAspLys	176
3620	ACAATGCGATGGAAAGAGCGGCTTAAGCTAGCACACAGCCGCTATTGATTAA	3669
177	GlnPheGlyThrGluAlaIalIysValGluSerAlaIalysAlaValThr..192	
3670	GCATCAACTCCAGATGCGACTTCAAGCTGAAGAGACAAAGGTGTACACCC	3719
193ThrGlnIysValGluValLys...199	
3720	TATCAATTGATTACTCTCCAAAGCAGATGCTTAAAGTGTCATGTCGTGA	3769
200PheSer	201
3770	AGTTGGCGATGAAATCAAGACAGCTCGAAGATTAAGCAAGCAGAAAGCAGAA	3819
202	LysAlaValGlu...LysLeuThrIysGluAspIleLysValThrAs	216
3820	AAACCTATTCGATGCGTCAACTATGACTCAATATGACAGAGAAAGCAATGCTAA	3869
216	AlYsAlaAsnAspIysValLeuValLysGluValIThrLeuSerGluA	233
3870	CAAGCCTCTTCACATGATTTTGTACATTAAAGAGAAAGACGAGCTTTGAAGACG	3919

233 sPLysArgSerAlaThrValGluLeuTyr..... 242
 243SerAsnLeuAlaAlaTyrGlnThrTyrThValAspValAs 256
 3920 CAGCTAGGGGTGACCAACAATGATCTTCATCAAGCTACTACAGAAAAA 3968
 3970 GCGAAGCGGGGGAACCTTGGTGGGAAAAAGCTTACA...GACACAGG 4016
 256 nLysValGluYsthrGluValAlaValGlySerLeuGlnAlaTyrTrhI 273
 4017 TAAAGAACCTAGAGATGCAATGCAATTAAATTG.....GCTAAGATTA 4054
 273 LeuMetAlaAspGlnThrValValAlaAspGluProthrAlaLeuGln 289
 4055 AAGAATTAGCTTAAGAGAACATCCGAAACAGAAAGAGAACACTTAATAA 4104
 290 PheThrValLys..... 293
 4105 ATAGTAGGAACCTTGCAGAAAGATACGCGCAAAAGCTATCGAGAACATCC 4154
 294AspGluAsnGly.....ThrGluVal 300
 4155 AAACCTGTCAAGATGAGATAGACAGCGAAATTAAAAAGCTAACTGACG 4204
 300 aLValSerPro.....GluGlyLeGluPheValTrh 310
 4205 CTGGCGCAAAACCTTAGACACACATTCGAGCAATGACAGATTAACGTAACG 4254
 311 ProAlaGluGlyLysLeuAsnAlaLysGlyGluLeuThrAlaLysGlu 327
 4255 CAGAGACACAAAAAAGCTAAGCGCCTAGCGATCTTGGAAAAAGCAATAAGA 4304
 327 YTrhSerTrhThrValLys.....AlaValYrL 337
 4305 AACACACAAAAATTGCGATAAAGCTGCGATTGATAGTTGACTTACTTCTG 4354
 337 YstLysAspGlyLysValAlaAlaGluSerLysGluValLys..... 350
 4355 TGAAGATGGTGGCTTGAAGCTACTTAAACAAGATGCTAAGAACACAGATT 4404
 351ValSerAlaGluGlyAlaAlaValAlaSerLysSerAsnTr 364
 4405 GCTAAAGATGACAGCCGCTCTAAAGAACCAATTCGACGCAATCCAAACTT 4454
 364 pThrValAlaGluGlnAsnLysAlaAspPheTrh.....SerL 377
 4455 GACAGACGCGAGG.....AAGAAAACTTCAACCGATGCGGTAGATGCGAG 4498
 377 yAspPheLysGlnAsnAsn.....Lys 384
 4499 AAGTACGCCAAACCTTAAGACGCAATTCACGTGCAACCCAGCCAGCAAGT 4548
 385 ValTyrGluGlyAspAsnAlaTyrValGlnValGluLeuLysAspGlnp 401
 4549 GTTCAAAAAGAAAGAGATGCGAGTGTTCGACCAATTGCAAGAAAGATGTCT 4598
 401 eAsnAlaVal.....ThrTrhLysLysValGluTyrGluSerLeuAla 415
 4599 TACGCGAGCTAAACAAGATGCTTAAGATTAAGATTGCTTAAGATGACGCG 4648
 415 snThrGluValAlaVal..... 420
 4649 CTGCTTAAGAGAACAAATTGCTCCAAATCCAAACTTGCACAGCCAGAGAG 4698
 421ValAspLysAlaThrGlyLysValTrh...Va 430
 4699 AAAACCTTCACCGATGCGGTAGATGACGAAGTGGCCAAAGCTTAAGACAGC 4748
 430 IleuSerAlaGlyLysAlaProValLysValTrhValLysAspSerLysG 447
 4749 AATTTCACGCTCAACAGCCACAGAGATGTTCAAAAGACAGAGATGACAG 4798
 447 LylLysAlaLeuValSerTrhTrhValGluTrhLeuAlaPheAlaGlnLys 463

4799 GTGTTGCGCCATTGCGAAGATGTT...CTTGACGCGAGCTAAACAGAT 4845
464 AlamelAspIleLysLeuGluLysThrAsnValAlaLeuSerThrLy 480
4846 GCTAAGATTAAGATGCTTAAGAA...TCCGACGCTGCTAACTCAGCCAT 4892
480 sAspValThrAspLeuLysValLysAlaProValLeuAspGlnTyrGlyL 497
4893 TGACGCGAAT...CCAAACTTGACACATCGACAGAGA 4924
497 ySgIuPheThr.....AlaProValThrValLys 506
4925 AGGAATCAAGCTTAAGAACAGTGTGATGCTAAAGCTCGACAGAT 4974
507 ValLeuAspLysAspGlyLysGluLeuLysGluGlnLysLeuGluAla 523
4975 GCAATGTGATGCTTCAACAAGTCCAGTCCAGGCAATCGCGACAGACAA 5024
523 sTyrValAsn.....ArgGluLeuValLeuAsnAlaAlaGlyGlnGlu 538
5025 AGCGCTAAGTTCATCGCCCAAGATGTTCTTGACGACGAAACAAGATG 5074
538 lagIyAsnTyrThrValValLeuThrAlaLysSerGlyGluLysGluAla 554
5075 CTAAGAAC.....AAGATTGCCAAAGAGTGTGC 5103
555 LysAlaThrLeuAlaLeuGluLeuLysAlaProGlyAlaPheSerLysph 571
5104 GCAGCTTAAGAACATGATGCTCAATCG..... 5133
571 eGIuValArgGlyLeuAspThrGluLeuAspLysTyrValThrGluGlu 588
5134 AACTTATCAGATGCGAGAACAGACTTCTAAGAACGCGTAGATGCG 5182
588 sngIuLys.....AsnAlaMetThrValSerValLeuProValAsp 601
5183 ATGCTTAAGCTACGACAGATGCAATGATGCTTCAACAAGTCAAGTGC 5232
602 AlaAsnGlyLeuValLeuLysGlyAla..... 610
5233 CGCGCATCGCGCAGAGACAAAGCGTAGTTCATCGCCCAAGATGTTCT 5282
611 GLuAlaAlaGluLeuLysValThrThrAsnLysGluGlyLysGluV 627
5283 TGACGCGACG.....AAACAAGATGCTTAAGATTAAGATTGCTAAAG 5326
627 alAspAlaThrAspAlaGlnValThrValGlnAsnAspSerValIleThr 643
5327 CCGACGCTGCTAAGTCAAGCCATTGACGCGAATCCAACTTGACAGATGCA 5376
644 ValGlyGlnGlyAlaLysAlaGlyGluThrTyrLysValThrValAla 660
5377 GAGAGAGATCGACTAG..... 5394
660 uAspGlyLysLeuIleThrThrHisSerPheLysValValAspThrAlaP 677
5395AAGCGGTAGAT..... 5406
677 roThrAlaLysGlyLeuAlaValaGluPheThrSerThrSerLeuLysGlu 693
5406 5406
694 ValAlaProAsnAlaAspLeuLysAlaAlaLeuLeuAsnIleLeuSerVa 710
5407GCAGATGCTAAAGCTGCG.....AC 5426
710 lAspGlyValProAlaThrThrAlaLysAlaThrAlaSerAsnValaGlu. 726
5427 AGATGCAATGTGATGCTTCAACAAGTCCAGTCCGACGCAATCGCGACAG 5476
727PheValSerAlaAsp..... 731

5477 ACAAGCGCTAGCCCATCGCCAAAGACATTTGATGCGCCGAAACA 5526
732ThrAsnValValaGlnAsnGlyThrValaGlyAlaLysGlyAl 746
5527 GATGCTAAGAACAGATGCTTAAGAG.....GCAGAAATCCGC 5564
746 aThrSerIleTyrValLysAsnLeuThrValValLysAspGlyLysGlu. 762
5565 TAAGTCAGCATGTGACTCCAAATCCGAACTTGACAGATGCGACTAAGAG 5614
763 ..GlnLysValaGluPheAspLysAlaValaGlnValaLaval 775
5615 CGCCTAAATCTGAATGTGCTTAAGCTGTGAGAACCGCAT 5655
seq_name: /cgn2/6/plodata/2/ina/6B_COMB.seq:US-08-913-942-3
seq_documentation_block:
Sequence 3, Application US/08913942
Patent No. 6200578
GENERAL INFORMATION:
APPLICANT: St. Geme, Joseph
APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr Hochbach Test Albritton & Herbert LLP
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913,942
FILING DATE: 29-DEC-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,995
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/4031
FILING DATE: 22-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Vance, Dolly A.
REGISTRATION NUMBER: 39,054
REFERENCE/DOCKET NUMBER: A-61053-1/RPT/RMS/DAV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 7291 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 163..7221
US-08-913-942-3
alignment_scores:
Quality: 240.50 Length: 850
Ratio: 0.584 Gaps: 41
Percent Similarity: 48.471 Percent Identity: 22.706
alignment_block:

US-09-754-947-1 x US-08-913-942-3 ..
Align seg 1/1 to: US-08-913-942-3 from: 1 to: 7291

```

23 GluysgLyAlaValLysGlyLysAspLysGlyMetPheGluProGlyLy 39
   |||||
397 GAAAGAGATTACAGAAATTCAAATTCAGGAAATATTTCGACAAATA 446
   |||||
39 sGluLeuThrArgAlaGluAlaAlaThrMetLeuLagInLeuAsnL 56
   |||||
447 AGGAGTACTAAAAAGCCGAGCAATCACCTCAAAAGCCGCGAC...AAC 493
   |||||
56 euProLLeaspLysAspAla.....LysProSerPheAla 67
   |||||
494 TGAAGATCAACAAACAAACCGATGAAAGCAACCAATCCAGTACTTCACC 543
   |||||
68 AspSerGlyGlyInTPtyrThrProPheIleAla...AlaValGly 83
   |||||
544 TACTCCCTGAAAAAGACCTCAACGATCTGACCAGTGTGCAACTGAAAA 593
   |||||
83 sAlaGlyVal.....IleLysGlyThrGlyVal 92
   |||||
594 ATTATGCTTTGGCGCAACGCGATAAAGTTGATATTACCAAGTATGCA 643
   |||||
92 sncGlyPheGlu.....ProAsnGlyLysIle..... 100
   |||||
644 ATGGCTTGAAATTTGGCGAAACACAGTAAACGGAATGTTTCATTGAA 693
   |||||
101 .....AspArgValSerMetAlaSerLeuLeuVal 111
   |||||
694 TTGGATTCAACTTTGCTGATGCGTACGAAATACAGGTGTGTTAACTTC 743
   |||||
111 uAlaIleTyrLysLeuAspThrLysValAsnGlyThrProAlaThrLysPhe 128
   |||||
744 ATCAAGTTTACACCTATATGATGTTGAAAAACAGAGCTGCACACTGTA 793
   |||||
128 yAspLeuGluThrLeuAsnTrpGlyLysGlyLysAlaAsnIleLeuVal 144
   |||||
794 AAGATGTTTAAATGCAAGTTGGAACATTAAGAGTGCTAAACAT..... 837
   |||||
145 GluLeuGlyIleSerValGlyThrGlyAspGlnTrpGluProLysLysTh 161
   |||||
838 .....GCTGGAGGTAAATGTTGAGAGTGTGATTT 866
   |||||
161 rValThrLysAlaGluAlaAlaGlnPheIleAlaLysThrAspLysGln 178
   |||||
867 AGTGTCCGCTTATATATATGTTGAATTATATACAGCGCATTAACACGCG 916
   |||||
178 heGlyThrGluAlaAlaLysValGluSerAlaLysAlaValThrThrGln 194
   |||||
917 TTGATGTTGTTTAAACAGCTAAAGAAAACGGTAAACCA..... 954
   |||||
195 LysValGluValLysPheSerLysAlaValAluLysLysLeuThrLysGluAs 211
   |||||
955 ...ACCGAAGTGAATTCACACCGAAACCTCTGTATCAAAAGAAAAGA 1001
   |||||
211 pIleLysVal...ThrAsnLysAlaAsnAsp.....LysValLeuVal 225
   |||||
1002 CGTAACTATTATTACTCGAATAAGCAATACACACAATTAAGTTACAA 1051
   |||||
225 aLysGluValThrLeuSerGluAspLysArgSerAlaThrValGluLeu 241
   |||||
1052 GTAAACACCGCATCTGATATACAGATGAGGGTAAATGGCTTAGTCACTGCA 1101
   |||||
242 TyrSerAsnLeuAlaAlaLysGlnThrTyrThrValAspValAspLysVal 258
   |||||
1102 AAAGCTGTGATTGATGCT.....GTCAACAAGGC 1130
   |||||
258 lGly.....LysThrGluValAla..... 264
   |||||
1131 TGCTTGAGAGATTAAACAACTACTGTAATGTCAAAATGCGCACTTCG 1180
   |||||
265 .....ValGlySerLeuGluAlaLysThrIleGluMetAlaAspGlnThr 279

```

```

1181 CAACGTGTCGTCAGCACAAATGTAACTTGAAGATGGCATGTGTACA 1230
   |||||
280 ValValAlaAspGluProThrAlaLeuGlnPheThrValLysAspGlu 296
   |||||
1231 ACAGCGTCA.....GTAATAAGATACTTA 1256
   |||||
296 nGlyThrGluValAlaSerPro.....GluGlyIleG 307
   |||||
1257 CCGCAATGGCATTCACACTGTAAGACAGCCGAAAGTTGGCAGCGCTTGA 1306
   |||||
307 LuPheValThrProAlaAlaGluLysIleAsnAlaLysGlyGluIleThr 323
   |||||
1307 AATTT.....GATACGATAAAAATC..... 1329
   |||||
324 LeuAlaLysGlyThrSerThrThrValLysAlaValTyrLysLysAspG 340
   |||||
1330 GTTCGCGATACGACCCGCACTTACTGTGACAGCT.....G 1364
   |||||
340 yLysValAlaAlaGluSerLysGlu.....ValLysValSerAlaG 354
   |||||
1365 TAAGTAGTGTGAATTCATAAGAGATGACAAAGAAAACTTTAATG 1414
   |||||
354 LuGlyAlaAlaValAlaSerIleSerAsn.....TrpThrValAlaGlu 368
   |||||
1415 CAGCGCAATTTGTTACACACTTTAGTATCTAGTGTGAAAGCAAAAGCT 1464
   |||||
369 GluAsnLysAlaAspPheThrSerLysAspPheLysGlnAsnLysVal 385
   |||||
1465 GAGCGTATCTGATGATGCGCTTGAGGAGATTTCAAAGAACCAAGAA 1514
   |||||
385 lTyrGluGlyAspAsnAlaTyrValGluValGluLeuLysAspGlnPhe 402
   |||||
1515 CAAGCGCGCAACCGGTA.....ACCTTTA 1540
   |||||
402 sAlaValAlaThrThrGlyLysValGluTyrGluSerLeuAsnThrGluVal 418
   |||||
1541 AAGCGGCGCAAGAACTTAAAGTGAACAGATGCGCAACTTACTTAT 1590
   |||||
419 AlaValAlaAspLysAlaThrGlyLysValThrVal..... 430
   |||||
1591 TCACGTGCAGATGCTTAAACGGGTTTAAACGAGCTTCTTAGTGGTAC 1640
   |||||
431 LeuSerAlaGlyLysAlaProValLysValThrValLysAsp..... 444
   |||||
1641 AACTAAATGGCGGAATGATGCGAAACCGTCAACAAACCGGTTAA 1690
   |||||
445 .....SerLysGlyLysAlaLeuValSerHisThrValGlu 456
   |||||
1691 CCATCAGCGCCAGAGTAAATGCGGTACGACAGGTACAAACCATTCAGC 1740
   |||||
457 lIleGluAlaPheAlaGlnLysAla...MetLysAspIleLysLeuGlu 472
   |||||
1741 GTAAACCAAGATGGCATTAAGCAGGTATTAAGCTATT..... 1779
   |||||
472 sThrAsnValAlaLeuSerThrLysAspValThrAspLeuLysValLysA 489
   |||||
1780 .ACTAATTTGCGAGTGTGTTAAGGCTTATGACGAT..... 1815
   |||||
489 lAProValLeuAspGlnTyrGlyLysGluPheThrAlaProValThrVal 505
   |||||
1816 .....GCGAATTTT 1824
   |||||
506 LysValLeuAspLysAspGlyLysGluLeuLysGluGln..... 518
   |||||
1825 GATGTTTAAATAACTGTGCACTCATTTAAATACACACGTGAAGATGC 1874
   |||||
519 .....LysLeuGluAlaLysTyrValAsnArgGlu...L 529
   |||||
1875 TTATAAGGTTTATTAATCTAAATGAAAGAAAATGCAATTAACAAACGT 1924
   |||||
529 euValLeuAsnAlaAlaGlyGlnGluAlaGlyAsnTyrThrVal..... 543
   |||||

```

```

1925 TGCGACTGACGACGCGCGGACTGAGCGCATTTACGTAATGGCT 1974
544 ..ValLeuThraLysSerGlyLysGluAa.....LysAlaThr 557
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1975 TGGGTACTATCAACCAAAAGGTACGAAAGCAATCAAGTTAA 2024
557 rLeuAlaLeuGluLeuLysAlaProGlyAlaPheSerLysPheGluValA 574
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2025 ACAAGCTGACGAAAGTCTTTACCGGGGCC..... 2055
574 rGlyLeuAspThrGluLeuAspLysThrValThrGluGluAsnGluLys 590
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2056 ..GGTGCTGCTACGGTTACTTCCAA.....TCGAAACGGTAA 2094
591 AsnAlaMetThrValSerValLeuProValAspAlaAsnGlyLeuValLe 607
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2095 CATACGATTACCGTTAGTGCGCGAAACGTAACGGGATTCGGCTCTGA 2144
607 uLysGlyAlaGluAlaGluLeuLysValThrThrThrAsnLysGluG 624
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2145 AAAAGATGGCGCATCTATTAGCTCAAGTGGATTAATCAAAAC..... 2187
624 LysGluValAspAlaThrAspAlaGluValThrValGluAsnAsnSer 640
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2188 .....ACTGATATGTTTAACTGTTGGTAATAT... 2217
641 ValIleThrValGlyGlnGlyAlaLysAlaGlyLutThrTyLysValTh 657
2217 ..... 2217
657 rValValLeuAspGlyLysLeuIleThrHisSerPheLysValValA 674
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2218 .....GTACTGCTGCACAAAGTCGCTTAAACCTGTTA 2254
674 sPThrAlaProThrAlaLysGlyLeuAlaValGluPheThrSerThrSer 690
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2255 AAATCGAGCGCACT.....GATGACAGATCGGGTAAGTAACACT 2292
691 LeuLysGluValAlaProAsn..AlaAspLeuLysAlaAlaLeuLeuAs 706
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2293 GTAAAGATGCTACTGCTAATGACGCTGATGAAGAGTCGCAACCTGTAA 2342
706 nIleLeuSerValAspGlyValProIleThrThrAlaLysAlaThrAla 723
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2343 AGATGTT.....GCACCGCAATTAAATAGTCGGCGGA 2374
723 eAsnValGluPheValSerAlaAspThrAsnValAlaIleGluAsnGly 739
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2375 CTTTGTGCAAAACAGAGAATTAACTACTCTATTGATGAAGATATCTCT 2424
740 ThrValGlyAlaLys.....GlyAlaThrSerIleTy 750
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2425 ACAGATTAACGCAAAAGATGACGCACTTAACGGCGATACCTTAACCTT 2474
750 rVal.....LysAsnLeuThrValValLysAspGlyLysGluGlnLys 765
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2475 TAAAGACAGGTAAACCTGAAAGTTAAAGTAGGAAAA.....AATA 2518
765 alGluPheLysAlaValGluValAlaValSerIleLysGluLysAlaLys 781
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2519 TTACTTTTGACTTGCGCAAAAC.....CTTGAAGTAAACGTAACGCAAA 2562

```

seq_name: /c9n2_6/plodata/2/lna/5A_COMB.seq:US-08-409-995-3

seq_documentation_block:

; Sequence 3, Application US/08409995

; Patent No. 5646259

; GENERAL INFORMATION:

; APPLICANT: Barenkamp, Stephen I.

; APPLICANT: St. Geme III, Joseph W.

; TITLE OF INVENTION: Haemophilus Adhesion Proteins

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

```

; ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/409,995
; FILING DATE: 24-MAR-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053/REF
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ. ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5738 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; US-08-409-995-3

```

```

alignment_scores:
  Quality: 238.00      Length: 850
  Ratio: 0.580        Gaps: 42
  Percent Similarity: 48.235  Percent Identity: 22.824

```

alignment_block:

US-09-754-947-1 x US-08-409-995-3 ..

Align seg 1/1 to: US-08-409-995-3 from: 1 to: 5738

```

12 HistPrgLysIleAspSerIleAsnTyrlleuValGluLysGlyAlaVally 28
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
256 AATTGGGAAATA.....TATTGACAAATAAAGAGTACTGTAA 293
28 s.....GlyAsnAspLysGlyMetPheGluP 37
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
294 AGCCGGACATCAACCCCTCAAAAGCCGGCAGACACCTGAAATMAACAAA 343
37 rGlyLysGluLeuThrArgAlaGluAlaIleThr.....MetAla 51
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
344 ANACGATGAAGAAAGNACATGCGAGTACCTTACCTACTCTGCAAAAAA 393
52 GlnIleLeuAsnLeuProIleAspLysAspAlaLysProSerPheAlaAs 68
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
394 GACCTCACAGATCTGACCAAGTGTTCGAACTGAAAAATTATGTGTGGCGC 443
68 pSerGlnGlyGlnTrpTyThrProPheIleAlaIleValGluLysAlaG 85
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
444 AAACGGCATTAA.....CTTGATTTTACACGATGATGCA 478
85 lValIleLys.....GlyThrGlyAsnGly...PheGluProAsnGly 98
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
479 ATGCTTGAAATTTGGCGAAACAGCTAACGGAATGTTCAATTGAAATGCT 528
99 .....LysIleAspArgValSerMetAlaSerLeuValGlu 111
529 TTGATTCACTTTCCTGATGCGGATGCAAGATACAGGATGTTAAGTTC 578
111 uAlaTyrlLysLeuAspThrLysValAsnGlyThrProIleThrLysPheL 128
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
579 ATCAAGTTTACACCTATGATGTTGAAAAAACAAGACACTGCAACTGTA 628

```

[illegible]

```
|||||
2090 AAGCTGAGCGACT.....GATGAGATCGCGGTAAGTAACT 2127
691 LeuylsGluValAlaProAsn...AlaSplLeuYsAlaAlaLeuLas 706
:::|||||
2128 GTAAAGATGCTACTGCTAATGACGCTGATGAAGACGCACTGTAA 2177
706 nileuSerValaspGlyValProIaThrThAlaLysAlaThrAlas 723
:::
2178 AGAGTT.....GCAACCGCAATTAATAGTCGGCGA 2209
723 eirsnValGluPheValSerAlaSplhrasnValAlaGluAsnGly 739
:::|||||
2210 CTTTGTGAAACAGAGAAATTAACTACTCTATGATGAGAAATACCT 2259
740 ThrValGlyAlaLys.....GlyAlaThrSerIleTy 750
||| ::::|||
2260 ACAGATACGCGCAAGATGACGCACTTAACGGCGCATACCTTAACCT 2309
750 rVal.....LysAsnLeuThrValValLysAspGlyLysGluLysv 765
:::|||||
2310 TTAAGCAGGTAAAAACCTGAAGTTAAACGTGATGAAAA.....ATA 2353
765 aGluPheAspLysAlaValGluValAlaValSerIleLysGluValLys 781
:::|||||
2354 TTTACTTTGACTTTGGCNAAAAAC.....CTTGAGGTGAAAAATGCGCAA 2397
```

seq_name: /cgn2_6/ptodata/2/lna/6A_COMB.seq:us-08-685-467-3

```
seq_documentation block:
: Sequence 3, Application us/08685467
: Patent No. 6060059
: GENERAL INFORMATION:
: APPLICANT: St. Geme III, Joseph W.
: APPLICANT: Barenkamp, Stephen J.
: TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Flehr, Hobach, Test, Albritton & Herbert
: STREET: Four Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: California
: COUNTRY: United States
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/685,467
: FILING DATE: 22-JUL-1996
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/409,995
: FILING DATE: 24-MAR-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Silva, Robin M.
: REFERENCE NUMBER: 38, 304
: REGISTRATION NUMBER: A-61053-2/RFT/RMS
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 781-1989
: TELEFAX: (415) 398-3249
: TELEX: 910 277299
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5738 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: unknown
: TOPOLOGY: unknown
: MOLECULE TYPE: DNA (genomic)
: US-08-685-467-3
```

```
alignment_scores:
Quality: 238.00 Length: 850
Ratio: 0.580 Gaps: 42
Percent Similarity: 48.235 Percent Identity: 22.824

alignment_block:
us-09-754-947-1 x us-08-685-467-3 ..
Align seq 1/1 to: us-08-685-467-3 from: 1 to: 5738

12 HistrrpGlyIleaspSerIleAsnTyrlLeuValGluLysGluValVal 28
:::|||||
256 AATTGGGGAATA.....TATTTCACATATAAGAGTACTTAA 293
28 s.....GlyAsnAspGlyGlyMetPheGlu 37
| ::::
294 AGCGGAGCAATACCTCAAGACCGCGACACCTGCAAAATNAAACAA 343
37 roGlyLysGluLeuThrArgAlaGluAlaThr.....MetMeta 51
||| ||:::|||||
344 ANACCGATGAAGNACCAATGCCAGTAGCTTACCTACTGCTGAAAAA 393
52 GluIleLeuAsnLeuProIleAspLysAspAlaLysProSerPheAlas 68
:::
394 GACCTCACAGATCTGACCAAGTGTGCAACTGAAAAAATTATCTTTGGCGC 443
68 pserGluGlyGlnTrpTyThrProPheIleAlaValGluLysLag 85
:::
444 AAACGGCGATTA.....GTTGATATTACCGATATGCA 478
85 lValIleLys.....GlyThrGlyAsnGly...PheGluProAsnGly 98
:::|||||
479 ATGCTTGAATTCGCGAAAAACGTAACGAAATGTTCAATTGAATGCT 528
99 .....LysIleAspArgValSerMetAlaSerLeuLeuValG 111
||| ||:::
529 TTGGATTCACTTTCCTGATGCGGTACGATACGATACAGGTGTTAAGTTC 578
111 uAlaTyrlLysLeuAspThrLysValAsnGlyThrProAlaThrLysPhe 128
:::
579 ATCAAGTTTACACCTAATGATGTTGAAGAAAAACAAGCTCACTGCTGA 628
128 yAspLeuGluThrLeuAsnTrpGlyLysGluLysAlaAsnIleLeuVal 144
|||||
629 AAGATGTTTAAATGACAGTTGAACATTAAGGTGCTAAACT..... 672
145 GluLeuGlyIleSerValGlyThrGlyAspGluTrpGluProLysLys 161
:::|||||
673 .....GCTGGAGGTAACTGTGAGCTGTGATTT 701
161 rValThrLysAlaGluAlaAlaGlnPheIleAlaLysThrAspLysGln 178
||:::
702 AGTCCGCTTATATATATGTTGAATTTATTACAGCGGATAAAAACAGC 751
178 heGlyThrGluAlaAlaLysValGluSerAlaLysAlaValThrGln 194
:::
752 TTGATGTTGTATTAAACAGCTAAAGAAAAACNGTAAACA..... 789
195 LysValGluValLysPheSerLysAlaValGluLysLeuThrLysGluAs 211
:::|||||
790 ...ACGAGAGTAATTCACACCGAAGAACCTCTGTTATCAAGAAAAAGA 836
211 pIleLysVal...ThrAsnLysAlaAsnAsp.....LysValLeu 225
| ||:::
837 CGGTAACTTATTTCTGGAAGAGATTAACGACACAAATAACTTCA 886
225 allYsGluValThrLeuSerGluAspLysArgSerAlaThrValGluLeu 241
:::
887 GTAACAGCGGAGCTGATATATACAGATGAGGTATAGCTTACACTGCA 936
242 TyrSerAsnLeuAlaAlaLysGlnThrTyThrValAspValAsnLysVa 258
::: |||
```

```

937 AAAGCTGATGATGCT.....GTGAACAAGC 965
258 LGly.....LysThrGluValAla..... 264
    ||| |||||
966 TGGTTGGACGACTTAACAACACTACTGCTAATGCAAAATGGCGACTTCG 1015
265 .....ValGlySerLeuGluAlaLysThrIleGluMetAlaAspGlnThr 279
    ||||| ||||| ||||| |||||
1016 CAACTGTTCCCTCAGGCACAATGTAACCTTTGAAAGTGGCGATGCTACA 1065
280 ValValAlaAspGluProThrAlaLeuGlnPheThrValLysAspGluAs 296
    ACACGGTCA.....GTAACATAAGATACTAA 1091
1066
296 nglyThrGluValAlaSerPro.....GluGlyIleG 307
    ||||| ||||| ||||| |||||
1092 CGCGAATGGCATCTCATCTTAAGTACGACGCCAAAGTTGGCGACGGCTTGA 1141
307 LurPheValThrProAlaAlaGluLysIleAsnAlaLysGlyGluIleThr 323
    ||||| ||||| ||||| |||||
1142 AATTT.....GATACGATAAAAAATC..... 1164
324 LeuAlaLysGlyThrSerThrThrValLysAlaValIleTyrLysAspGlu 340
    ||||| ||||| ||||| |||||
1165 GTTCACATACAGACCGCAGCTTACTGTACAGCT.....GG 1199
340 YLysValValAlaGluSerLysGlu.....ValLysValSerAlaG 354
    ||||| ||||| ||||| |||||
1200 TAAAGTAGCTGAATAATCTTAAGAAATGACAAAGAAACCTTGTTATG 1249
354 LucGlyAlaAlaValAlaSerIleSerAsn.....TrpThrValAlaGlu 368
    ||||| ||||| ||||| |||||
1250 CAGCGCATTTGGTAAACAGCTTTAGTAATCTAAGTTGAAAGCAAAAGCT 1299
369 GlnAsnLysAlaAspPheThrSerLysAspPheLysGlnAsnLysVal 385
    ||||| ||||| ||||| |||||
1300 GAGCGTGATCTCATGTGCGCTTGAAGGATTTCAAAAGCAACAGAGAGT 1349
385 IYrGluGlyAspAsnAlaTyrValGluValGluLeuLysAspGlnPheA 402
    ||||| ||||| ||||| |||||
1350 CAAAGCAGCGAAACGTA.....ACCTTTA 1375
402 snAlaValThrThrGlyLysValGluTyrGluSerLeuAsnThrGluVal 418
    ||||| ||||| ||||| |||||
1376 AAGCGGCAAGAACTTAAGGTAAACAGAGATGTCGAACTTTACTTAT 1425
419 AlaValAlaAspLysAlaThrGlyLysValThrVal..... 430
    ||||| ||||| ||||| |||||
1426 TCACCTCAGATGCTTAAACGGGTTAACGACATTACTTTAGCTGTAC 1475
431 LeuSerAlaGlyLysAlaProValLysValThrValLysAsp..... 444
    ||||| ||||| ||||| |||||
1476 AACTAATGCGGAAATGATGCGAAACCGTCATCAACAAAGCGTTTAA 1525
445 .....SerLysGlyLysAlaLeuValSerIleThrValGlu 456
    ||||| ||||| ||||| |||||
1526 CCATCAGCGCAGCAGGTATGGCGGTACGACAGCTCAAAACACCATCAGC 1575
457 IleGluAlaPheAlaGlnLysAla..MetLysAspIleLysLeuGluIly 472
    ||||| ||||| ||||| |||||
1576 GTAAACCAAGATGCGATTAAACAGAGTAAAGCTATT..... 1614
472 sThrAsnValAlaLeuSerThrLysAspValThrAspLeuLysValLysA 489
    ||||| ||||| ||||| |||||
1615 ACTAATGTCGAGTGGGTTAAGAGCTTATGACGAT..... 1650
489 laProValLeuAspGlnTyrGlyLysGluPheThrAlaProValThrVal 505
    ||||| ||||| ||||| |||||
1651 .....GCGAATTTT 1659
506 LysValLleuAspLysAspGlyLysGluLeuLysGluGln..... 518
    ||||| ||||| ||||| |||||
1660 GATGTTTAAATACTGCAACGATTTAAATAGACACAGCTTGAAGATGC 1709

```

```

519 .....LysLeuGluAlaLysTyrValAlaAsnArgGlu...L 529
    ||||| ||||| ||||| |||||
1710 TTATAAAGCTTTATTAATCTTAATGAAAAAATGCAATTAACAAACCGT 1759
529 euValLleuAsnAlaAlaGlyGlnGluAlaGlyAsnTyrThrVal..... 543
    ||||| ||||| ||||| |||||
1760 TGGTGACTGACAGCAGCGGCGGACTGTAGCGATTACGTAATTTGGGT 1809
544 ..ValLeuThrAlaLysSerGlyLysGluAla.....LysAlaThr 557
    ||||| ||||| ||||| |||||
1810 TGGGTAGATCAACCAAAACGATACGAAAGAAAGCAATCAAGTTAA 1859
557 rLeuAlaLeuGluLeuLysAlaProGlyAlaPheSerLysPheGluValA 574
    ||||| ||||| ||||| |||||
1860 ACAAGCTGATGAAGTCTCTTTACCGGAGCC..... 1890
574 rGlyLeuAspThrGluLeuAspLysTyrValThrGluLysGlnLys 590
    ||||| ||||| ||||| |||||
1891 ..GGTGTGCTACGCTTACTTCAAA.....TCTGAACCGTAA 1929
591 AsnAlaMetThrValSerValLeuProValAspAlaAsnGlyLeuValle 607
    ||||| ||||| ||||| |||||
1930 CATACGATTACCGTTAGTGGCTGAACAACTAAACGGATTCGGCTTGA 1979
607 uLysGlyAlaGluAlaAlaGluLeuLysValThrThrThrAsnLysGluG 624
    ||||| ||||| ||||| |||||
1980 AAAAGATGGCGATTAATTAAGCTCAAGTGATTAATCAAAAC..... 2022
624 LyrGluValaAspAlaThrAspAlaGlnValThrValGlnAsnSer 640
    ||||| ||||| ||||| |||||
2023 .....ACTGATAATGTTTAACTGTGTGTAATAT... 2052
641 ValIleThrValGlyGlnGlyAlaLysAlaGlyGluThrTyrLysValTh 657
    ||||| ||||| ||||| |||||
2052 ..... 2052
657 rValValLeuAspGlyLysLeuIleThrThrHisSerPheLysValValA 674
    ||||| ||||| ||||| |||||
2053 .....GCTACTGTCTACTTAAGGTGGCTTGAACCTGTTA 2089
674 sPThrAlaProThrAlaLysGlyLeuAlaValGluPheThrSerThrSer 690
    ||||| ||||| ||||| |||||
2090 AAACGTGAGCGACT.....GATGCAGATCGCGGTAAAGTAAC 2127
691 LeuLysGluValAlaProAsn..AlaAspLeuLysAlaAlaLeuLeuAs 706
    ||||| ||||| ||||| |||||
2128 GTAAAGATGCTACTCTTAATGACGCTGATTAAGAAAGTCGCAACTGTAA 2177
706 nIleLeuSerValAspGlyValProAlaThrThrAlaLysAlaThrAlas 723
    ||||| ||||| ||||| |||||
2178 AGATGTT.....GCAACCGCAATTATATAGCGCGCA 2209
723 eraAsnValGluPheValSerAlaAspThrAsnValAlaAlaGluAsnGly 739
    ||||| ||||| ||||| |||||
2210 CTTTGTGTAACAGAAATTAACCTCTTATTCATGATGATATTCCT 2259
740 ThrValGlyAlaLys.....GlyAlaThrSerIleTyr 750
    ||||| ||||| ||||| |||||
2260 ACAGATTAACGGCAAGATACGACACTTAAAGCGGGGATACCTTAACCTT 2309
750 rVal.....LysAsnLeuThrValValLysAspGlyLysGluGlnLysV 765
    ||||| ||||| ||||| |||||
2310 TAAAGCAGGTAAAAACCTGAAGTTAAACGTATGAGAAA.....ATA 2353
765 aGluPheAspLysAlaValGlnValAlaValSerIleLysGluAlaLys 781
    ||||| ||||| ||||| |||||
2354 TTACTTTGACTTGGCNAAAAAAC.....CTTGAGGTGAAAACTCGAAA 2397

```

```

seq_name: /cgn2_6/prodata/2/lna/5A_COMB.seq:US-08-038-682-8
seq_documentation_block:
; Sequence 8, Application US/08038682

```



```

439 LysValThrValLysAspSerLysGlyLysAlaLeuValSerHisThrVa 455
|||
4425 AAAGTGGAAATTGAATCACTTCGGTATATGTAATATTACAGCGAGCGG 3474
445 LgluileGluAlaPheAlaGlnLysAlaMetLysAspIleLysLeuGluL 472
|||
3475 CAATACACTTAAGTAAGTAATATCAGTGTGTCAGATCTAACAGTAACAG 3524
472 yStrAsnValAlaLeuSerThrLysAspValThrAspLeuLysValys 488
|||
3525 CGGATGCAGAGCCCTGCACAACTACAGCAGGCTCAACCATTAAGCGACA 3574
489 AlaProValLeuAspGlnIleGlyLysGluPheThrAlaProValThrVa 505
|||
3575 ACAGGC.....AATGCAATATTACAC 3597
505 LysValLeuAspLysAspGlyLysGluLeuLysGluLysLeuGluA 522
|||
3598 CAAACAGCTGATATCAACGCT.....AAAGTTGAAT 3629
522 LgluLysValAlaSerGluLeuValLeuAsnAlaAlaGlyGluGluAla 538
|||
3630 CCAGCTCCGCTGTGTAACCTGTGTAACCTGAGCAGCACTCTGCTGTA 3679
539 GlysAsnTyr.....ThrValValLeuThrAlaLysSerGlyGlyLys 552
|||
3680 GGTATATTTCAGTACACTGTTACTATTACTGCGGATAGCGGT..... 3724
552 sGluAlaLysAlaThrLeuAlaLeuGluLeuLysAlaProGlyAlaPheS 569
|||
3725 .....AAATTAACCTCCACAGTAGTGTTCATCAATTAATGGACTAATA 3767
569 eLysPheGluValAlaGlyLeuAspThrGluLeuAspLysTyrValThr 585
|||
3768 GT.....CTAAC 3775
586 GluLysAsnGlnLys.....AsnAlaMetThrVa 595
|||
3776 ACCTCAACCAATCAGCGGATATTGAGGTACAAATTTCTGTAATACAGT 3825
595 lSerValLeuProValAspAlaAsnGlyLeuValLeuLysGlyAlaGluA 612
|||
3826 AAATGTT...ACAGCAAGCAGCTGATTTAATCTTGAATTAAGTACGCA 3872
612 lAlaGluLeuLys.....ValThrThrAsnLysGluLys 625
|||
3873 AAGTTGAAGCGGAAATGAGCTGCAACCTTAAGTGAATGAGGCAAA 3922
626 GluValAspAlaThrAspAlaGlnValThrValGlnAsn...AsnSerVa 641
|||
3923 TTAACCCACCAACAGGCTGTAGCATTCCTCAACCAATGTCAGACAAAC 3972
641 lIleThrValGlyGlnGlyAlaLysAlaGlyLysThrLysValThrVa 658
|||
3973 TCTTACACCCAAAGATGACAGATATGCGAGAAACATTAATGCTGCTAATG 4022
658 alValLeuAsp.....GlyLysLeuIleThrThrHisSerPheLysVal 672
|||
4023 TGAGCTTAAATTAACACAGGCACTTAACACTACAGAGGATTCAG... 4069
673 ValAspThrAlaProThrAlaLysGlyLeuAlaValGluPheThrSerTh 689
|||
4070 .....ATTAAACGCAACCAAGTGG 4086
689 rSerLeuLysGluValAlaProAsnAlaAspLeuLysAlaAla..... 703
|||
4087 TACCTTAACATCAATGCAAAAGATGCCAAATTAATGATGCTGCATCAG 4136
704 .....LeuLeuAsnIleLeuSerValAspGlyValProAlaThr 716
|||
4137 GTGACCGCAGCACTAGTAAATGCACTAACGCAAGTGGCTGTGTAACGIG 4186

```

```

717 ThrAlaLysAlaThrAlaSerAsnValGluPheValSerAlaAsp..... 731
|||||
4187 ACTGGGAAA...ACCTCAAGCAGCGTGAAT...ATCAGCGGGAATTTAA 4230
732 .....ThrAsnValAlaAlaGluAsnGly.....ThrValG 742
|||||
4231 CACATTAATGCGTTAAATATCATTTCCGAAATGCGTGAACACTGTC 4280
742 LysAlaLysGlyAlaThrSerIleTyrValLysAsnLeuThrValValys 758
|||
4281 CCTTAAGAGGC...AAGCAATGTGATGTGAAA.....TATATCCAA 4318
759 AspGlyLysGluGlnLysValGluPheAspLysAlaValGlnValAlaVa 775
|||
4319 CCAGGTGAGCAAGCGGTGAAGAGGTAATGAGCGAAACGCGTCCTTGA 4368
775 lSerIleLysGlu 779
|||||
4369 GAGGTAAAGAT 4381

seq_name: /cgn2_6/prodata/2/lna/5A_COMB.seq:us-08-302-832-8
seq_documentation_block:
; Sequence 8, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,832
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US pct/us93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4702 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-302-832-8

alignment_scores:
Quality: 218.50 Length: 771
Ratio: 0.530 Gaps: 36

```


[illegible]

```

2746 CCAAAATGGCGGCAATATCTCACAACAAAAGGCAATCTCACAATTTCTT 2795
255 aAaLnLySValGly.....LysThrGluValAla 264
   ::|||::|||::|||::|||::|||::|||
2796 CCGATAAATAATATTTACCAATCAGATTAACAATCAACAAAGGCGGTGA 2845
265 ValGlySerLeuGluValAlaThrIleGluMetLAspGlnThrVala 281
   |||::|||::|||::|||::|||::|||
2846 GGGGGCGGCTTCGATTCAGATTCAGGACAAATAATCTCAACCTAATATCA 2895
281 lAlaSpGlu...ProThrAlaLeuGlnPheThrValLysAspGluSng 297
   ::|||::|||::|||::|||::|||::|||
2896 AACCAAGAGGTTAAAAATTGGCAGAGAGACTTAATATTTCCAGCCTTATA 2945
297 LYThrGluValAlaSerProGluGlyIleGluPheValThrProVala 313
   ::|||::|||::|||::|||::|||::|||
2946 AAGCAAAATTCACCTAATAAATGGCAGCTGATTTACATTTGGCAATGCT 2995
314 GluLYsIleAsnAlaLysGly...GluIlePheuAlaLysGlyThrSe 329
   ::|||::|||::|||::|||::|||::|||
2996 AGCGGTGATATGCTGATCTGATTAATAAAGTGACTTTTGCAAGGTTAAAG 3045
329 rGThrValLYsAlaValTYrLYsLYsAspGlyLYsValValaLAsLus 346
   ::|||::|||::|||::|||::|||::|||
3046 TTCAAAATCTCGACT.....GAGGTCACATGTACACTAA 3083
346 erLYsGluValLYsValSerAlaGluGlyAlaAlaValAlaSerIleSer 362
   ::|||::|||::|||::|||::|||::|||
3084 ATAGCCAAATGTAAACGCTCT...AATGGTAGTACGATGCGGTAAATGAT 3130
363 AsnTYrThrValAlaGluGlnAsnLYsAlaAspPheThrSerLYsAspPh 379
   ||| ||| ||| ||| ||| |||
3131 AACAGACGCGGTTTAACATTTCCGCAAAAGATGTAACGGTAACAACTAA 3180
379 eLYsGlnAsnAsnLYsValTYrGluGlyAspAsnAlaTYrValGlnValG 396
   ::|||::|||::|||::|||::|||::|||
3181 CGTTACTCTCCACACAGACATTAATATCTCGCCGACACAGAAATGTAA 3230
396 lueLYsAsp.....GlnPheAsnAlaValThrThrGlyLYsValGlu 410
   ::|||::|||::|||::|||::|||::|||
3231 CAACCAAAAGGACACAACTATCAATGCA..ACCACGAGCGGTGGA 3277
411 TYrGluSerLeuAsn..... 415
   ::|||::|||::|||::|||::|||::|||
3278 GTAACTGCTCAAAATGCTCAATTAAGGCAACATTACTCGCAAAATGT 3327
416 .....ThcGluValAlaValVala 422
   ||||| ||||| ||||| |||||
3328 AACAGTGCACAGACACGAAAAATCTGTTCACACAGACATCTGTCACTTA 3377
422 sPLysAlaThrGlyLYsValThrValLeuSerAlaGlyLYsAlaProVal 438
   ::|||::|||::|||::|||::|||::|||
3378 ATGCACACCGCGCACAGTAATCACTT...AGTACAAAACAGGGGATATT 3424
439 LYsValThrValLYsAspSerLYsGlyLYsAlaLeuValSerThrIstIrra 455
   ::|||::|||::|||::|||::|||::|||
3425 AAAGGTGAATTCGAATCACTTCGCTGATATGTAATTAATTAACGCGACG 3474
455 lGluIleGluAlaPheAlaGlnLYsAlaMetLYsAspIleLYsLeuGluL 472
   ::|||::|||::|||::|||::|||::|||
3475 CAATACACTTAAGTAAGTAATATCATCAGTCGTCACAGATTAACACTAACG 3524
472 ySThrAsnValAlaLeuSerThrLYsAspValThrAspLeuLYsValLYs 488
   ::|||::|||::|||::|||::|||::|||
3525 CGGATCGAGAGCGCTTGACAACTACAGACGAGCTCAACCATTAAGTCGACA 3574
489 AlAProValLeuAspGlnTYrGlyLYsGluPheThrAlaProValIrra 505
   ::|||::|||::|||::|||::|||::|||
3575 ACAGGC.....AATGCAAATATTTCACAC 3597
505 lLYsValLeuAspLYsAspGlyLYsGluLeuLYsGluGlnLYsLeuGluA 522
   ::|||::|||::|||::|||::|||::|||
3598 CAAAAACAGGATATCAACGGT.....AAATTTGAT 3629

```

522	laLystyValaIasnArgLueValLeuIasnAlaIaIagIngInguaIa	538
523	539
524	540
525	541
526	542
527	543
528	544
529	545
530	546
531	547
532	548
533	549
534	550
535	551
536	552
537	553
538	554
539	555
540	556
541	557
542	558
543	559
544	560
545	561
546	562
547	563
548	564
549	565
550	566
551	567
552	568
553	569
554	570
555	571
556	572
557	573
558	574
559	575
560	576
561	577
562	578
563	579
564	580
565	581
566	582
567	583
568	584
569	585
570	586
571	587
572	588
573	589
574	590
575	591
576	592
577	593
578	594
579	595
580	596
581	597
582	598
583	599
584	600
585	601
586	602
587	603
588	604
589	605
590	606
591	607
592	608
593	609
594	610
595	611
596	612
597	613
598	614
599	615
600	616
601	617
602	618
603	619
604	620
605	621
606	622
607	623
608	624
609	625
610	626
611	627
612	628
613	629
614	630
615	631
616	632
617	633
618	634
619	635
620	636
621	637
622	638
623	639
624	640
625	641
626	642
627	643
628	644
629	645
630	646
631	647
632	648
633	649
634	650
635	651
636	652
637	653
638	654
639	655
640	656
641	657
642	658
643	659
644	660
645	661
646	662
647	663
648	664
649	665
650	666
651	667
652	668
653	669
654	670
655	671
656	672
657	673
658	674
659	675
660	676
661	677
662	678
663	679
664	680
665	681
666	682
667	683
668	684
669	685
670	686
671	687
672	688
673	689
674	690
675	691
676	692
677	693
678		


```

411 TyrGluSerLeuAsn..... 415
    :|||
3278 GTAACTGCTCAAAATGCTACAAATTAAAGCAACATTACCTCGCAAAATGCT 3327
416 .....ThrGluValAlaValAla 422
3328 AACAGTGCACGCAACAGAAATCTTTGTTACCAAGAGATGCTGTCATTA 3377
422 splysAlaThrGlyLysValThrValLeuSerAlaGlyLysAlaProVal 438
3378 ATGCACACGAGGCGCACTAAACATT...ACTACAAAACAGGGGATTTT 3424
439 LysValThrValLysAspSerGlyLysAlaLeuValSerHisThrVa 455
3425 AAAGGTGAATTGAATCAACTCCGGTATGTAATATTATACAGCAGCGG 3474
455 IGLIILeGluAlaPheAlaGlnLysAlaMetLysAspIleLysLeuGluL 472
3475 CAATACACTTAAGTAACTAAATATCAGTCTGTCAGATGTAAACGTAACAG 3524
472 yStrAsnValAlaLeuSerThrLysAspValThrAspLeuLysValLys 488
3525 CGGATGCGAGAGCCTTGACAACTACAGCAGGCTCACCATTAAGTCCGACA 3574
489 AlaProValLeuAspGlnTyrGlyLysGluPheThrAlaProValThrVa 505
3575 ACAGGC.....AATGCAAAATATTACAAAC 3597
505 LLysValLeuAspLysAspGlyLysGluLeuLysGluGlnLysLeuGlu 522
3598 CAATAACAGCTATATCAACGGT.....AAAGTTGAT 3629
522 LAlsTyrValAsnArgGluLeuValLeuAsnAlaAlaGlyGlnGluAla 538
3630 CCAGTCCGGCTCTGTACACTTGTCGCACTGAGCAACACTCTGCTGTA 3679
539 GlyAsnTyr.....ThrValValLeuThrAlaLysSerGlyGlyLys 552
3680 GGTATATATTTCAGGTACGTAACCTGTTACTGCGGATAGCGGT.... 3724
552 sGluAlaLysAlaThrLeuAlaLeuGluLeuLysAlaProGluAlaPheS 569
3725 .....AAATTAACCTCCACAGTAGGTTCTACATTATAGGACATATA 3767
569 erLysPheGluValArgGlyLeuAspThrGluLeuAspLysTyrValThr 585
3768 GT.....GTAAAC 3775
586 GluGluAsnGlnLys.....AsnAlaMetThrVa 595
3776 ACCTCACAGCCATCAGGCAATTCGAAGTACAAATTCTGTAATACAGT 3825
595 lSerValLeuProValAspAlaAsnGlyLeuValLeuLysGlyAlaGlu 612
3826 AAATGTT...ACAGCAGCAGCTGGGATTTACATTATGGAATATAGTCGCA 3872
612 lAlaGluLeuLys.....ValThrThrThrAsnLysGluGlyLys 625
3873 AAGTTGAAGCAAAATAGAGCTGCACTTAACGTCGATCAGCAAGCAAA 3922
626 GluValAspAlaThrAspAlaGlnValThrValGlnAsn...AsnSerVa 641
3923 TTAACCAACCAACAGAGCTTACGATTCCTCAAGCAATGTCAGACAAAC 3972
641 lIleThrValGlyGlnGlyAlaLysAlaGlyGluThrTyrLysValThrV 658
3973 TCTTACAGCCAAAGATAGCAGTATCGCAGAACTTAATCTGCTAATG 4022
658 alValLeuAsp.....GlyLysLeuIleThrThrHisSerPheLysVal 672
4023 TGAGGTTAAATACCAAGCAGCACTTAACACTACAGGGGATTTCAAAG... 4069

```

```

673 ValAspThrAlaProThrAlaLysGlyLeuAlaValGluPheThrSerTh 689
4070 .....ATTACGCCAACCGATGG 4086
689 rSerLeuLysGluValAlaProAsnAlaAspLeuLysAlaAla..... 703
4087 TACCTTACAAATCAATGCAAAAGATGCCAAATTAGATGCTGCTGCATCAG 4136
704 .....LeuLeuAsnIleLeuSerValAspGlyValProAlaThr 716
4137 GTGACCCGACAGTAGTAATATCACTAACGCAAGAGGCTGTGTAACGTG 4186
717 ThrAlaLysAlaThrAlaSerAsnValGluPheValSerAlaAsp.... 731
4187 ACTGGAAA...ACCTCAAGCAGCGTGAAT...ATCACCAGGGGATTTAAA 4230
732 .....ThrAsnValValAlaGluAsnGly.....ThrValG 742
4231 CACAAATAATGGTTAAATATCATTTTCGAAAATGTTAGAAACACTGTGC 4280
742 lValAlaLysGlyAlaThrSerIleTyrValLysAsnLeuThrValValLys 758
4281 GCTTAAGAGC...AAGAAATTGATGTGAAA.....TATATCCAA 4318
759 AspGlyLysGluGlnLysValGluPheAspLysAlaValGlnValAlaVa 775
4319 CCAGGTGTAGCAAGCGTGAAGAGGTAATGAAAGCAACGCGTCTTGA 4368
775 lSerIleLysGlu 779
4369 GAAGGTAAAGAT 4381

seq_name: /cgn2_6/plodata/2/lna/SB_COMB.seq:us-08-728-470-8
seq_documentation block:
; Sequence 8, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-633
; TELECOMMUNICATION INFORMATION:

```

```

:
: TELEPHONE: (703) 415-0810
:
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4702 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-728-470-8

alignment_scores:
    Quality: 218.50      Length: 771
    Ratio: 0.530          Gaps: 36
    Percent Similarity: 53.437    Percent Identity: 22.438

alignment_block:
US-09-754-947-1 x US-08-728-470-8 ..
Align seg 1/1 to: US-08-728-470-8 from: 1 to: 4702

89 GlyThrGlyAsnGlyPheGluProAsnGlyLysIleAspArgValSerIle 105
   |||
   |||
2306 GGGGAAATTTCAAGCAGTAGCATTCAGGGCAATATCAATATCCACCAATTA 2355
105 LAlaSerLeuLeuValGluAlaIleTyrLysLeuAspThrLysValAsnGlyT 122
   |||
   |||
2356 ACCAATCTTACATTACAGCT.....GACACGACACACAGCAACA 2396
122 hPrcAlaThrLysPheLysAspLeuGluThrLeuAsnTrpGlyGlu 138
   |||
   |||
2397 CAGGCTTCAGGAAA.....AGAACTCAACTCTGGC..... 2428
139 LysAlaAsnIleLeuValGluLeuGlyIleSerValGlyThrGlyAspGlu 155
   |||
   |||
2429 .....AATATATCTGTGGGCAATTTTAAGCTTA...ACTGGTGCACA 2469
155 nTrpGluProLysLysThrValThrLysAlaGluAlaIleGlnPheIleA 172
   |||
   |||
2470 TGCAGAACATTTGCGGCAATCTTCTATTCAGCAA..... 2503
172 LaLysThrAspLysGlnPheGlyThrGluAlaIleLysValGluSerAla 188
   |||
   |||
2504 .....GATTCCACATTTAAAGAGAGCCAGTGACACCACTAAACATC 2545
189 LysAlaValThrThrGlnLys.....ValGluValLysPheSerLysAl 203
   |||
   |||
2546 ACCGGCACCTTTACCAACACAGCGTACCGCCACATTATATTAACACAGG 2595
203 aValGluLysLeuThrLysGluAspIleLys.....V 214
   |||
   |||
2596 ACTGGTAAACATCCAAAGCGCATATTATCAATAAGGTGTTAAATATCA 2645
214 aLThrAsnLysAlaAsnAsnAspLysValLeuValLys..... 226
   |||
   |||
2646 CTACTAACGCCCTCAGGCACTCAAAAAACCATTTAAACGGAATATACT 2695
227 .....GluValThrLeuSerGluAspLysArgSerAlaThrVa 239
   |||
   |||
2696 AACGAAAAAGGCGACTTAACATCAAGATATTAAAGCCGACGCGCAAT 2745
239 GluLeuTyrSerAsnLeuAlaIleLysGln...ThrTyrThrValAspV 255
   |||
   |||
2746 CCAATTTGGCGGCATATCTCAGAAAAAGAACGCAATCTCACAATTTCT 2795
255 aLAsnLysValGly.....LysThrGluValAla 264
   |||
   |||
2796 CTGATTAAGCTAAATATTACCAATTCAGATTAACATCAAGACGCGTTGA 2845
265 ValGlySerLeuGluAlaIleLysThrIleGluMetAlaAspGlnThrValVa 281
   |||
   |||
2846 GCGGGCCCTTCTGATTCAAGTGAGCGCAAAATGCTAACCTAATTAATCA 2895
```

```

281 LAlaAspGlu...ProThrAlaLeuGlnPheThrValLysAspGluAsnG 297
   |||
   |||
2896 AACCAAGAGTTAAATTTGGCAGAGAGACTTAATATTTAGGCTTAAATA 2945
297 LThrGluValValSerProGluGlyIleGluPheValThrProAlaAla 313
   |||
   |||
2946 AAGCAGAAATTCACGCTAAAAATGGCAGATTTAACTATTTGGCAATGCT 2995
314 GluLysIleAsnAlaLysGly...GluIleThrLeuAlaLysGlyThrSe 329
   |||
   |||
2996 AGCGGTGTAATGCTGATGCTAAAAAGTAGCTTTTGACACAGGTTAAAGA 3045
329 rThrThrValLysAlaValIleTyrLysAspGlyLysValAlaIleGlu 346
   |||
   |||
3046 TTCAAAAATCTCGACT.....GACGCTCACAATGTAAACACTTA 3083
346 eLysGluValLysValSerAlaGluGlyAlaAlaValAlaSerIleSer 362
   |||
   |||
3084 ATAGCGAAGTGAAGAACGTCT...AATGTATGACCAATGCTGTATGAT 3130
363 AsnTrpThrValAlaGluGlnAsnLysAlaAspPheThrSerLysAspPh 379
   |||
   |||
3131 AACAGCACCGGTTTAACCATTTCCGCAAAAGATGTACGTAACCAATTA 3180
379 eLysGlnAsnAsnLysValIleTyrGluGlyAspAsnAlaTyrValGlnValG 396
   |||
   |||
3181 CGTTACCTCCCAAGACACATTAATATCTCTGCCGACAGCAAGAAATGTA 3230
396 LuLeuLysAsp.....GlnPheAsnAlaValThrThrGlyLysValGlu 410
   |||
   |||
3231 CAACCAAAAGAACGACCAACTCAATGCA...ACCACAGCGACGCTGGAA 3277
411 TyrGluSerLeuAsn..... 415
3278 GTAACCTGCTCAAAATGTGTAATTAAGCAACATTACTCCCAAAATGT 3327
416 .....ThrGluValAlaValAla 422
   |||
   |||
3328 AACAGTGACAGACACAGAAATCTTGTACACAGAGAAATGCTGCATTA 3377
422 sPLysAlaThrGlyLysValThrValLeuSerAlaGlyLysAlaProVal 438
   |||
   |||
3378 ATGCAACGACGCGCACAGTAAACATT...AGTACAAAACAGGAGATAT 3424
439 LysValThrValLysAspSerLysGlyLysAlaLeuValSerHisThrVa 455
   |||
   |||
3425 AAAAGTGCAATGAAATCAACTTCGCTAATGTAAATATTACGCGAGCGG 3474
455 GluIleGluAlaPheAlaGlnLysAlaMetLysAspIleLysLeuGluL 472
   |||
   |||
3475 CAATACACTTAAGTAAGTAATATATCATCTGTCACAGATGTAAACAGTAACAG 3524
472 ystThrAsnValAlaLeuSerThrLysAspValThrAspLeuLysValLys 488
   |||
   |||
3525 CGGATGCAAGACCTTGAACACTACAGCAGGCTCAACCACTTATGTCGACA 3574
489 AlaProValIleuAspGlnTyrGlyLysGluPheThrAlaProValThrVa 505
   |||
   |||
3575 ACAGGC.....AATGCCAAATATTACAAAC 3597
505 LlysValLeuAspLysAspGlyLysGluLeuLysGluGlnLysLeuGluA 522
   |||
   |||
3598 CAAAACAGGTGATATCAACGGT.....AAAGTTGAAAT 3629
522 LaLysTyrValAsnArgGluLeuValLeuAsnAlaIleGlyGlnGluAla 538
   |||
   |||
3630 CCAGCTCGGGCTCTGTAAACACTTGTGCAAACGTGACCAACTCTCTGCTGA 3679
539 GlyAsnTyr.....ThrValValLeuThrAlaLysSerGlyGlyLys 552
   |||
   |||
3680 GSTAATATTTCAGTAAACAGCTGTACTATTACTGCGGATAGCGGT..... 3724
```



```

4137 GTGACGCACAGTAGTAATGCAACATGACCAAGTGGCTCTGGTAACGTG 4186
||||| :||||| :||||| :|||
717 ThrAlaLysAlaThrAlaSerAsnValGluPheValaSerAlaasp.... 731
||||| :||||| :||||| :|||||
4187 ACTGCGAAA...ACCTCAACACACGCTGAAT...ATCACCGGAGATTAA 4230
||||| :||||| :||||| :|||||
732 .....ThrAsnValValAlaGluAsnGly.....ThrValG 742
||||| :||||| :||||| :|||||
4231 CACAATTAATGGTGAATATTCATTTCGGAAAATGGTAACACATGTCG 4280
742 ValAlaLysGlyAlaThrSerIleTyValLysAsnLeuThrValaLys 758
:||||| :||||| :||||| :|||||
4281 GCTTAAGAGCG...AAGCAATATGATGTGAA.....TATATCAA 4318
759 AspGlyLysGluLysValGluPheAspLysAlaValaGluValaLava 775
||| :||| :||| :|||
4319 CCAGGTGTGACGAACGCTAGCAAGAGGTAAATTGAAGCGAAACGCGTCTGA 4368
775 lSerIleLysGlu 779
:||||| :||||| :||||| :|||||
4369 GAAGGTAAAAAGAT 4381

seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-617-697-8

seq_documentation_block:
; Sequence 8, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: Of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4803 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-617-697-8
alignment_scores:

```

Quality:	218.50	Length:	771
Ratio:	0.530	Gaps:	36
Percent Similarity:	53.437	Percent Identity:	22.438

Align seg 1/1 to: US-08-617-697-8 from: 1 to: 4803

89	gltYtHglyAsnGlyPheGluIuProAsnGlyLysIleAspArgValSerIse	105
2521	GGGGAAATATTCAGACAGTATTCAGGGCAATATCAATATACCAATATA	2570
105	TAAsSerLeuLeuValAGlualATyLysLeuAspThrLysValasnGlyT	1122
2571	ACCAAAATGTTACATTTACAAAGCT.....GACACCAAGCAACAGCAACA	2611
122	hPProAlaThrLysPheLysAspLeuGluThrLeuAsnTrpGlyLysGlu	138
2612	CAGGCTTCAGAAATA.....AGAACTCTAACCTCTGGC.....	2643
139	LysAlaAsnIleLeuValGluLeuGlyIleSerValGlyThrGlyAspGly	155
2644AAATATCTGTTGGAGGGGAATTTAAGCCCTA...ACTGGTCGAAA	2684
155	nTrpGluPProLysThrValThrLysAlaGluAlaIaGlnPheIleA	172
2685	TGCAAACATGTCGGGCATCTTTCTATTCAGAAA.....	2718
172	LaLysThrAspLysGlnPheGlyThrGluAlaAlaLysValGluSerAla	188
2719GATTCACATTTTAAAGAGAGAGCCAGCAGTCAACCTAAATC	2760
189	LysAlaValThrThrGlnLys.....ValGluValLysPheSerLysAl	203
2761	ACCGGACACTTTATCCACAAACAGCGTACCCCAACATTAATATAAACAGG	2810
203	avaAlGluLysLeuThrLysGluAspIleLys.....V	214
2811	AGTGGTAATACCTCCAAAGGAGATATTATCAATAAAGGTGGTTAATATCA	2860
214	aLThrAsnLysAlaAsnAsnAspLysValLeuValLys.....	226
2861	CTACATACGCCCTCAGGCACTCAAAAAACCATTAATTAAAGCAAAATATACT	2910
227GluValThrIleuSerLusAspLysArgSerAlaThrIva	239
2911	AACGAAAAAGCGCACTTAAACATCAAGATATTTAAAGCCGAGCGCAAT	2960
239	IgluLeuTyTyrSerAsnLeuAlaIaLysGln...ThyTyrThrValAspV	255
2961	CCAATATGGCGGCATATCTCCACAAAAGAGGCAATCTCACAATTTCTT	3010
255	aLAsnLysValGly.....LysThrGluValAla	264
3011	CTGATTAAGTAATATTATACCATCATGATTAACATCAAGCAAGCGGCTTGAA	3060
265	ValGlySerLeuGluAlaLysThrIleGluIleThrIleAspGlnThrVala	281
3061	GGGGGGCGTTCTGATTTCAAGTGAGGCAACAATAAGCTTAACCTAATATCA	3110
281	IaIaAspGlu...ProThrIaLeuGlnIuPheThrValLysAspGluAsnG	297
3111	AACCAAGAGGTTAAATATGGCAGGAGACCTTAATATTTCCAGGCTTTATA	3160
297	LyrThrGluValValSerProGluGlyIleGluIuPheValThrProAlaIa	313
3161	AAGCGAAATATTACACTTAAATATGCACTGATTTTAACATTATGGCAATGCT	3210
314	GluLysIleAsnAlaLysGly...GluIleThrLeuAlaLysGlyThrIse	329
3211	AGCGGTGGTAATGCTGATGATCAAAAAGATGACACTTTTGCAAGGTTAAAGA	3260


```

      /gene="sap"
      /note="surface layer homology motif, SLH1"
      mls_c-feature      866..1042
      /gene="sap"
      /note="surface layer homology motif, SLH2"
      mls_c-feature      1043..1216
      /gene="sap"
      /note="surface layer homology motif, SLH3"
BASE COUNT      1360 a      471 c      645 g      986 t
ORIGIN

alignment_scores:
      Quality: 3847.00      Length: 785
      Ratio: 4.919      Gaps: 0
      Percent Similarity: 99.618      Percent Identity: 98.981

alignment_block:
US-09-754-947-1 x BASAPG ..

Align seg 1/1 to: BASAPG from: 1 to: 3462

1 AlaglyLysThrPheProAspValProAlaAspHisTrpGlyIleAspSe 17
|||||
674 GCAGGTAAACATTCGCCAGAGCTTCCTGTCATCAGTGGGAAATTGATTC 723

17 rIleAspTrpLeuValGluLysGluValAlaValLysGluAsnAspLysGlu 34
|||||
724 TATTACTACTAGTAGAGAAAAGCGCAGTTAAAGTTAAACGACAAAGGAA 773

34 ePheGluProGlyLysGluLeuThrArgAlaGluAlaAlaThrMetMet 50
|||||
774 TGTTCAGACCTGGAAGAAATTAACCTCGCGAAGCAGCTACATGATG 823

51 AlagInIleLeuAsnLeuProIleAspLysAspAlaLysProSerPheAl 67
|||||
824 GCTCAATCTTAACTTACCAATCGATTAAGAGTCTAAACCATCTTTGCG 873

67 AspSerGlnGlyGlnTrpTrpThrProPheIleAlaAlaValGluLysA 84
|||||
874 TGACCTCTCAGGCCAATGTCACATCCATCTTCGACGCTGTAGAAAAG 923

84 lAglyValIleLysGlyThrGlyAsnGlyPheGluProAsnGlyLysIle 100
|||||
924 CTGGCGTTATTAAAGGTACAGGAAACGGCTTGAGCCAAACGAAAAATC 973

101 AspArgValSerMetAlaSerLeuLeuValGluAlaIleTrpLysLeuAsp 117
|||||
974 GACCCGCTTCTATGCGATCTCTCTGTGAGAACTTACAAAATTAGATAC 1023

117 rLysValAsnGlyThrProAlaThrLysPheLysAspLeuGluThrLeuA 134
|||||
1024 TAAACTAAACGCTACTCCGCAACTAAATTTCAAGATTTAGAAACATTTA 1073

134 sntTrgGlyLysGluLysAlaAsnIleLeuValGluLeuGlyIleSerVal 150
|||||
1074 ACTGGGGTAAAGAAAAGCTAAACATCTAGTGAATTAGGAATCTCTGTT 1123

151 GlyThrGlyAspGlnTrpGluProLysLysThrValIleLysAlaGluAl 167
|||||
1124 GGTACTCGTGATCAATGGAGCCTTAAGAAAACCTTAACCTAAAGCAGAAC 1173

167 AlaGlnPheIleAlaLysThrAspLysGlnPheGlyThrGluAlaAlaL 184
|||||
1174 TGCTCAATTCATTGCTAAGAGCTGACACAGCAGTTGCGTACAGAAAGCAGA 1223

184 ysvAlGlnSerAlaLysAlaValIleThrGlnLysValGluValLysPhe 200
|||||
1224 AAGTGAATCTGCAAAAGCTGTACAACTCAAAAAGTAAAGTTAAATTTC 1273

201 SerLysAlaValGluLysLeuThrLysGluAspIleLysValIleThrAsn 217
|||||
1274 ACCAAAGCTGTGAAAATTAACATAAGAGATATCAAAAGTAACTATACAA 1323

```

```

217 sAlaAsnAsnAspLysValIleuValLysGluValIleuSerGluAspL 234
|||||
1324 AGCTAACACAGATTAAGTACTAGTTAAAGAGCTACTTTATCAGAAAGTA 1373

234 ysrArgSerAlaThrValGluLeuTrpSerAsnLeuAlaAlaLysGlnThr 250
|||||
1374 AAAAATCTGTACAGTTGAATTATATTAAGTAACTTATAGCAGCTAAACAA 1423

251 TyrThrValAspValAsnLysValGlyLysThrGluValAlaValGlySe 267
|||||
1424 TACACTGTAGATGTAAACAAAGTTGTGTAACACAGAGTAGCTGAGTTTC 1473

267 rLeuGluAlaLysThrIleGluMetAlaAspGlnThrValAlaAlaAspG 284
|||||
1474 TTTAGAACCAAAACAAATCGAAATGCTGACCAAAACATTTAGCTGATG 1523

284 lUpProThrAlaLeuGlnPheThrValLysAspGluAsnGlyThrGluVal 300
|||||
1524 AGCCAAACAGCATTTCAATTCACAGTTAAAGATGAAGAACGTTACGAACTT 1573

301 ValSerProGluGlyIleGluPheValIleProAlaAlaGluLysIleAs 317
|||||
1574 GTTTCACCAAGAGGATTAATTTGTAAAGCCAGCTGCAGAAAAAATTAA 1623

317 nAlaLysGlyGluIleThrLeuAlaLysGlyThrSerThrValLysA 334
|||||
1624 TGCAAAAGGTGAATCACTTTAGCAAAAGGTCACTCACTACTGTAAAG 1673

334 lAvalIleLysLysAspGlyLysValValAlaGluSerLysGluValLys 350
|||||
1674 CTGTTTATAAAAAGACGTTAAAGTAGTAGCTGAAGGTAAAGATTAAGA 1723

351 ValSerAlaGluGlyAlaAlaValAlaSerIleSerAsnTrpThrValAl 367
|||||
1724 GTTTCGTGGAAGTGTGCTGCACTTACCTCAATCTTACTGACAGTTGCC 1773

367 aGluGlnAsnLysAlaAspPheThrSerLysAspPheLysGlnAsnAsnL 384
|||||
1774 AGAACAAATATAAGCTGACTTACTTCTAAAGATTCAAAACAAACATA 1823

384 ysvAlIleGluGlyLysAsnAlaIleValGluValLysAspGln 400
|||||
1824 AAGTTTACGAAAGGCAACAGCTTACGTCAAGTCAATGCAATTCAGATCAA 1873

401 PheAsnAlaValIleThrGlyLysValGluTrpGluSerLeuAsnThrGl 417
|||||
1874 TTTAACGCGAGTAAACAACAGTGAATATGAGTGGTTAAACACAGA 1923

417 uValAlaValAlaAspLysAlaThrGlyLysValIleThrValLeuSerAla 434
|||||
1924 AGTTGCTGTAGTAGTAAGCTACTGTAAGTAAGTAACTGTATTATGTCAG 1973

434 lLysAlaProValLysValIleThrValLysAspSerLysGlyLysAlaLeu 450
|||||
1974 GAAAAGCACCAATAAAGTAAGTAAAGATTTCAAAAGTAAAGAACTT 2023

451 ValSerHisThrValGluIleGluAlaPheAlaGlnLysAlaMetLysAs 467
|||||
2024 GTTTCAAAAACAGTTGAATTTGAGCTTCCGCTCAAAAAGCAATGAAGA 2073

467 lLysLysLeuGluLysThrAsnValAlaLeuSerThrLysAspValIleTh 484
|||||
2074 AATTAAATTAGAAAAAACTAACGTAGCGCTTCTCAAAAAGATGTAACAG 2123

484 sPheLysValLysAlaProValIleuAspGlnIleLysGluPheThr 500
|||||
2124 ATTTAAAGTAAAGAGCTCCAGTACTAGATCAATACGTTAAAGATTTACA 2173

501 AlaProValIleThrValLysValLeuAspLysAspGlyLysGluLeuLysGl 517
|||||
2174 GCTCCTGTAAAGTAGTAAGTACTGATAAAGATGTAAAGATTTAAAGA 2223

```

```
517 uGluLysLeuGluAlaLysTyrValAsnArgGluLeuValLeuAsnAla 534
|||||
2224 ACAAAATTAAGGAAATATATGTAAGCAAAAGATTAGTCTGTAATCAG 2273
534 IaGluGluAlaGlyAsnTyrThrValValLeuThrAlaLysSerGly 550
|||||
2274 CAGGTCAAGAAAGCTGTAATTAATACACTTGTATTAACTGCAAAATCGCT 2323
551 GluLysGluAlaLysAlaThrLeuAlaLeuGluLeuLysAlaProGlyAl 567
|||||
2324 GAAAAAGAACCAAAAGCTACATTAGCTCTGTAATTAAGACTCCAGGTGC 2373
567 aPheSerPheGluValArgGlyLeuAspThrGluLeuAspLysTyr 584
|||||
2374 ATTCTTAATTTGAAGTCTGCTTGAAGAAAAGAAATTGATTAATGATG 2423
584 aThrGluGluAsnGluLysAsnAlaMetThrValSerValLeuProVal 600
|||||
2424 TTACTGAGGAAACCAAAAGATGCAATGACTGTTTCAGTTCTTCTCTGTA 2473
601 AspAlaAsnGlyLeuValLeuLysGlyAlaGluAlaIaGluLeuLysVa 617
|||||
2474 GATGCAATAGATGATGATTAATAAGGTGCAGAAAGCAGCTGAATAAAGT 2523
617 LThrThrThrAsnLysGluGlyLysGluValAspAlaThrAspAlaGln 634
|||||
2524 AACACAAACAACCAAGAGAGTAAGAGAGAGAGAGAGAGAGAGAGAGAG 2573
634 aThrValGlnAsnAsnSerValIleThrValGlyGlnGlyAlaLysAla 650
|||||
2574 TTACTGACCAAAATTAACAGTGTATTACTGTTGCTCAAGGTGCAAAAGCT 2623
651 GlyLysThrTyrLysValIleThrValLeuAspGlyLysLeuIleThrTh 667
|||||
2624 GGGAACCTATAAAGTAACAGTGTACTGATGATGTAATTAATCACACAC 2673
667 rHisSerPheLysValValAspThrAlaProThrAlaLysGlyLeuAla 684
|||||
2674 TCATCTCTTAAGAGTGTGATACAGCACCAACGCTAAAGGATTACAG 2723
684 aLgIuPheThrSerThrSerLeuLysGluValAlaProAsnAlaAspLeu 700
|||||
2724 TAGAATTACAGCACATCTCTTAAGAGTACCTCCAAATGCTGATTTA 2773
701 LysAlaAlaLeuLeuAsnIleLeuSerValAspGlyValProAlaThrTh 717
|||||
2774 AAGCTCAGCTTTTAATATCTTAATCTGATGATGCTGACCTGCGACTAC 2823
717 rAlaLysAlaThrAlaSerAsnValGluPheValSerAlaAspThrAsn 734
|||||
2824 AGCAAAAGCAACAGTTTCTAATGATGAAATTGTTTCTGCTGACACAATG 2873
734 aValAlaIaGluAsnGlyThrValGlyValLysGlyAlaThrSerIleTyr 750
|||||
2874 TTGTAGCTGAAATGTTACAGTGTGTCACAAAGGTGCACAACTTCTAT 2923
751 ValLysAsnLeuThrValValLysAspGlyLysGluGlnLysValGluPh 767
|||||
2924 GTGAAAAAAGCTGACAGTGTAAAGAGATGCAAAAGCAAAAGTAGAATT 2973
767 eAspLysAlaValGlnValAlaValSerIleLysGluAlaLysProAla 784
|||||
2974 TGATTAAGCTGTACAAAGTTCAGTTCTAATTAAGAGCAAAAGCTGCAA 3023
784 hrlYs 785
|||||
3024 CAAAA 3028
seq_name: gp_ba:BRH012290
seq_documentation_block: 4082 bp DNA BCT 25-OCT-1998
LOCUS BRH012290
DEFINITION Bacillus thuringiensis etc gene.
```

```
ACCESSION AU012290
VERSION AU012290.1 GI:3805794
KEYWORDS ctc gene; CTC protein; S-layer protein.
SOURCE Bacillus thuringiensis.
ORGANISM Bacillus thuringiensis
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Bacillus; Bacillus cereus group.
1 (bases 1 to 4082)
REFERENCE
AUTHORS Sun,M.
TITLE Direct Submission
JOURNAL Submitted (22-OCT-1998) Sun M., Department of Microbiology,
Huazhong Agricultural University, Department of Microbiology,
Huazhong Agricultural University, Wuhan, Hubei 430070, CHINA
2 (bases 1 to 4082)
Sun,M. and Yu,Z.
S-layer protein gene of Bacillus thuringiensis CTC
Unpublished
FEATURES
source location/Qualifiers
1. 4082
/organism="Bacillus thuringiensis"
/sub_species="Finlilinus"
/strain="CTC"
/db_xref="taxon:1428"
/note="Flagellar serotype H2"
718. 3168
/gene="ctc"
718. 3168
/gene="ctc"
/function="S-layer protein"
/codon_start=1
/transl_table=11
/product="CTC protein"
/protein_id="CA09981.1"
/db_xref="GI:3805795"
/translation="MAKINSYKKVIAGTMTAAVAVVSPVAAGKSPDVPADHMG
DSINLVKGAATGNDKGMPEBGEKLTBAATMAOILNLPIDKADPSFADSGOM
YTPFAAEKAKGVITGNGPBNKIDRYSAKSLVBAKIDTPVNGKSTAYELSNLA
TLNMGKRNILVELGISVGTADWPERKTYTKAAQOFIAKTROGTEARAKESA
KAVTQKVEKFSKAVEVGLTEKEDVLANKANDKVLVDKLSDEKSAVELSNLA
AKOTYTVNKNKVEVEVGLTEKEDVLANKANDKVLVDKLSDEKSAVELSNLA
GIEFVTPAEKINAKGEITLAKGSTYKAVYKKGKVAESKEVVAEGTAVASIS
NMWVAERADFTSKDFKONDKVEGDNNVOVELKDOHNNVNNVNAKAESELENTEAV
VDKATGKTVLSAGKAPYKTVYKSKKEIYSKYEIAPAFQAKAKELIKETYNALS
TKDVIDEKVAPVLDQYGERAPAVEVVLVDGKELBOKLVAKYENKELVLANHG
BAGKYTVELTAASGRKEVSKLAELEKAPGVFSKDVGLLENLEDKVYTEENKEMV
VSLPVDANGVLREKEAATLKVTDDGKVDAISQVAVNDAAGTITVENKAKG
ETVKTVAADGKLITTHSEFKVVDTPAPAKKLAVDPTSLNEVAOGSEKTRALNLS
VDGVPATAGATVTDVTKFVSADTNVNSEETAKFGKGSIFVKELTVKKKGQKVE
LDKPYRVDSIYEKVEK"
BASE COUNT 1574 a 543 c 765 g 1200 t
ORIGIN
alignment_scores:
Quality: 3304.50 Length: 785
Ratio: 4.424 Gaps: 5
Percent Similarity: 95.159 Percent Identity: 86.497
alignment_block:
US-09-754-947-1 x BRH012290 ..
Align seg 1/1 to: BRH012290 from: 1 to: 4082
1 AlaGlyLysThrPheProAspValProAlaAspHisTrpGlyIleAsp 17
|||||
805 GCAGGAAATCATTTCCAGACGTTCCAGCTGATCAGCTGGGAATTGATTC 854
17 rIleAsnTyrLeuValGluLysGlyAlaValLysGlyAsnAspLysGly 34
|||||
855 TATTAAGCTTGTAGTAAAGGCGGCTTACAGTAACGATTAAGGAA 904
34 ePheGluProGlyLysGluLeuThrArgAlaGluAlaIaThrMetMet 50
|||||
```

```

905 TGTTCGAGCCCTGGAAAGAAATTAACTGCTGCAGAAAGCAGCTACATGATG 954
51 AlaGlnIleLeuAñLeuProIleAspLysAspAlaLysProSerPheAl 67
955 GCTCAAAATCTTAAACTTACCAATCGATTAAGATGCTTAACCATCTTCCG 1004
67 aasSerGlnGlyGlnTrpTyrThrProPheIleAlaIaValGluLysA 84
1005 TGACTCTCAAGGCCAATGTGACACTCCATTCATCCGACGCTGTAGAAAAAG 1054
84 laGlyAlaIleLysGlyThrGlyAsnGlyPheGluProAsnGlyLysIle 100
1055 CTGGCCTTATTAAAGGTACAGCAAGCGCTTTGAGCCAAAGCAAAATTC 1104
101 AspArgValSerMetAlaSerLeuLeuValGluAlaTyrLysLeuAspTh 117
1105 GACCGCGTTTCAATGGCATCTCTCTGTGAAAGCTTACAATTTGATATC 1154
117 rLysValAsnGlyThrProAlaThrLysPheLysAspLeuGluThrLeuA 134
1155 TAAAGTAAACGGTACTCCAGCACTAAATTCAAAGATTGTAGAAACATTAA 1204
134 snTrpGlyLysGluLysAlaAsnIleLeuValGluLeuGlyLysSerVal 150
1205 ACTGGCGTTAAGAAAAAGCTAACATCCAGTTGAATTAGGAATCTCTGTT 1254
1251 Gly...ThrGlyAspGlnTrpGluProLysLysThrValThrLysAlaG 166
1255 GGTACTACAGCTGATAGTGGACCTTACGAAACCTGTAAGCTAAAGCAGA 1304
166 uAlaIleGlnPheIleAlaLysThrAspLysGlnPheGlyThrGluAla 183
1305 AGCTGCTCAATTTATGCTTAAGACTGCAAAACAGTTGCGTAGCAAGACTAG 1354
183 laLysValIleLysSerAlaLysAlaValThrThrGlnLysValGluValLys 199
1355 CAAAAGTTGAATCGTAAAGCAGTTACACTCAAAAGCTAAGTAGAAGTTAA 1404
200 PheSerLysAlaValGluLysLeuThrLysGluAspIleLysValThrAs 216
1405 TTCAGCAAACTGCTGTAATAATTAACTAAAGAAAGACGTTAAATTTAGCTAA 1454
216 nLysAlaAsnAsnAspLysValLeuValLysGluValThrLeuSerGluA 233
1455 CAAAGCTAACAAACGATTAAGCTATTAGATCTTAAATTAATTAATCAGAG 1504
233 sPLysArgSerAlaThrValGluLeuTyrSerAsnLeuAlaAlaLysGln 249
1505 ACAAAAATCTGCTACAGTTGAATTAATAGTAAGTACTGCTGCTAACAA 1554
250 ThrTyrThrValAspValAsnLysValGlyLysThrGluValAlaValG 266
1555 ACTTACAGCTGATGATTAATTAAGCTTGGTAAAGTAGAGGTAACGTGAGG 1604
266 ySerLeuGluAlaLysThrIleGluMetAlaAspGlnThrValValAla 283
1605 TTCCTTAACCAACAAAACAATTTGAATGGCTGACCAAAACGCTGTACCTG 1654
283 sPGluProThrAlaLeuGlnPheThrValLysAspGluAsnGlyThrGlu 299
1655 ATGAGCCACATGCAATTAATTAACATTAAGATGAAGTGAAGCGGTAGCGAA 1704
300 ValValSerProGluGlyIleGluPheValThrProAlaIleGluLys 316
1705 GTTGCTTTACCAGCGCGTATTGAATTTGTACCTCCAGCTGCAGAAAAAT 1754
316 eaSnAlaLysGlyGluIleThrLeuAlaLysGlyThrSerThrThrVal 333
1755 TAATGCAAAAGGTGAATTAATTAACCAAAAGGTACTTCACTACGTGTA 1804
333 ysaLalaValTyrLysLysAspGlyLysValAlaValAlaGluSerLysGluVal 349
1805 AACGCTGTTATATAAAAAAGACGGTAAGCTAGCTGTAAGATGAAGAAAGTA 1854
350 LysValSerAlaGluGlyAlaAlaValAlaSerIleSerAsnTrpThrVal 366
1855 AAAGTTTCTGCTGAGAGTACTGACAGTAGCTTCAATCTTAACGACAGCT 1904
366 laIleGluGlnAsnLysAspPheThrSerLysAspPheLysGlnAsnA 383
1905 AGCAGCGAAG...AAAGTGATTTCACTTCTTAAAGACTTCAAAACAAAG 1951
383 snLysValTyrGluGlyAspAsnAlaTyrValGlnValGluLeuLysAsp 399
1952 ATTAAGTTTATGAAAGGTGACAAAGCTTCTGTCACAGTGAATTAAGAC 2001
400 GlnPheAsnAlaValThrThrGly...LysValGluTyrGluSerLeuAs 415
2002 CATTTCAATTAAGCTGTAAATTAACGTTAAAGCTGAATTAAGATTAATA 2051
415 nThrGluValAlaValAlaAspLysAlaThrGlyLysValThrValLeu 432
2052 TACAGAGGTAGCTGTAGATTAAGCTACGTGAAGTAACTGATATAT 2101
432 eAlaGlyLysAlaProValLysValThrValLysAspSerLysGlyLys 448
2102 CTGCAAGAAAAGCACCACTTAAGTAAGTGTAAAGATTCAAAAGGTAA 2151
449 AlaLeuValSerHisThrValGluIleGluAlaPheAlaGlnLysAla 465
2152 GAACCTGTTCAAAAACAGTGAATTAAGCTTTTCTCTCAAAACGCAAT 2201
465 LysAspIleLysLeuGluLysThrAsnValAlaLeuSerThrLysAsp 482
2202 GAAGAATTAATTAAGAAAACCTAAGCTAGCGCTTCTACAAAGATG 2251
482 aThrAspLeuLysValLysAlaProValLeuAspGlnTyrGlyLysGlu 498
2252 TAACAGATTTCAAAGTAAAGCTCCAGTCACTGATCAATATGCTAAAGAG 2301
499 PheThrAlaProValThrValLysValLeuAspLysAspGlyLysGlu 515
2302 TTGCGAGCTCTCTTAAGATGAAGTAAGTGAATTAAGATGCTAAAGAA 2351
515 uLysGluGlnLysLeuGluAlaLysTyrValAsnArgGluLeuValLeuA 532
2352 AAAAGAACAAAATTAAGTGAATTAAGTGAATTAAGATTAAGATTAAG 2401
532 snAlaIleGlyGlnGluAlaLysAlaThrLeuAlaLeuGluLeuLysAla 548
2402 ATGCACATGCGCAAGAGCTGTTAATTAATTAAGTGAATTAAGATTAAG 2451
549 SerGlyGlyLysGluAlaLysAlaThrLeuAlaLeuGluLeuLysAla 565
2452 TCTGTTAAAAAAGAGTGAATTAATTAAGTGAATTAAGATTAAGATTAAG 2501
565 oGlyIlePheSerLysPheGluValArgGlyLeuAspThrGluLeuAsp 582
2502 AGCGCTATTTCTTAATTAATTAAGTGAATTAAGTGAATTAAGATTAAG 2551
582 yTyrValThrGluGluAsnGlnLysAsnAlaMetThrValSerValLeu 598
2552 AATACGTTACTGAAGAAACAAAAGAACAAATGCGTAGTTCTGTTCTT 2601
599 ProValAspAlaAsnGlyLeuValLeuLysGlyAlaGluAlaLeuGlu 615
2602 CCACTAGATGCTAATGTTAGTTTAAAGAGAAAGAAAGACGACCAACT 2651
615 uLysValThrThrAsnLysGluGlyLysGluValAspAlaThrAspA 632
2652 TAAAGTACTACTACAGATTAAGATGTAAGTAGGATGACAACTTCAG 2701
632 laGlnValThrValGlnAsnAsnSer...ValIleThrValGlyGlnGly 647
2702 GTCAAGTACTGTAAATGATGCTGCTGGCACAAATTAAGTGAAGAAATGA 2751

```


648 AAlaAlaAlaGlyGluThrTyrLysValThrValValLeuAspGlyLysLe 664
|||||
2752 GCAAAGCTGGTGAAGAAATATACTAAGTGTAGTACAGACGCTAAGCT 2801
664 uIlleThrHisSerPheLysValAlaSpHrAlaProHrAlaLysG 681
|||||
2802 AATCCACACTCATTCATTTAAAGTTGTGACACAGCACCCTGCAGCTAANA 2851
681 lYLeuAlaValAlaGluPheHrSerThrSerLeuLysGluValAlaProAsn 697
|||||
2852 AATTAGACAGTGTAGATTTCACAGACATCTCTTAACGAGGTGCTCAAGGT 2901
698 AlaAspLeuLysAlaAlaLeuLeuAsnIlleLeuSerValAspGlyValPr 714
|||||
2902 AGCGAATTAATAAACTGCACCTCTTAATATCCATCATGTTGATGAGTGTCC 2951
714 oAlaThrThrAlaLysAlaThrAlaSerAsnValGluPheValSerAla 731
|||||
2952 AGCAACAACCTCGGGTGCACACTGACGTGACGTGAATTTGTCTGCAG 3001
731 sPThrAsnValAlaAla...GluAsnGlyThrValAlaGlyAlaLysGlyAla 746
|||||
3002 ATACAAATGTTGTAGTACAGAACTGCTAAGTGTGTACAAAGTCTCT 3051
747 ThrSerIleTyrValLysAsnLeuThrValValLysAspGlyLysGluG 763
|||||
3052 ACTCAATCTTGTGTAAGAACTGTTCTTAATAAAGGTGACGACAAACA 3101
763 nLysValGluPheAspLysAlaValGluValAlaValSerIleLysGluAl 780
|||||
3102 AAAAGTTAGCTTGACAAAGCAGTACAGTACGATGTTCTTATTAAGAAG 3151
780 lAlys 781
:::|
3152 TAAAA 3156
seq_name: gb_ba:BFH249446
seq_documentation_block:
LOCUS BFH249446 3159 bp DNA BCT 01-JUN-2001
DEFINITION Bacillus thuringiensis slpA gene for surface-layer protein.
ACCESSION AJ249446
VERSION AJ249446.1 GI:6580204
KEYWORDS slpA gene; surface-layer protein.
SOURCE Bacillus thuringiensis.
ORGANISM Bacillus thuringiensis
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/staphylococcus group; Bacillus; Bacillus cereus group.
REFERENCE
1 (bases 1 to 3159)
Ménage,S., Hauslant,M. and Pouet,A.
A general strategy for identification of S-layer genes in the
Bacillus cereus group: molecular characterization of such a gene in
Bacillus thuringiensis subsp. gallieriae NRRL 4045
Microbiology 147 (Pt 5), 1343-1351 (2001)
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
2 (bases 1 to 3159)
Ménage,S.
Direct Submission
Submitted (14-SEP-1999) Ménage S., Bacteriologie / Mycologie,
Institut Pasteur, 28 rue du Dr Roux, PARIS 75724 Cedex 15, FRANCE
FEATURES
Source
Location/Qualifiers
1. 3159
/organism="Bacillus thuringiensis"
/strain="NRRL 4045"
/sub_species="gallieriae"
/db_xref="taxon:1428"
270..2751
/gene="slpA"
270..275
/gene="slpA"
/note="contains 3 SLH (S-Layer Homology) motifs towards
its amino-terminus"
/evidence="experimental"

CDS
286..2751
/gene="slpA"
/function="unknown"
/codon_start=1
/transl_table=11
/evidence="experimental"
/product="surface-layer protein"
/protein_id="CAB63252.1"
/db_xref="GI:6580205"
/db_xref="SPRMBL:09RED0"
/translation="MAKNSYKKYVIAAGTAAVAVGVSPVAAGKTFPPDVGPSMA
EYDYLVAKAIKGRPDGTPPEALDRASAATIMATLELEVKDQAKPFRQADSM
AAKYLAVAEQAVIQDGTGNFNPSSNINASASAVTKVTKLEDRVSGELTFNSOL
KDHMGKRDANIILVALITNGTGMWEDKSVTRAEAKFTAKTDMQFGQAEAKVSI
KEINAKIEYKFGTEVDYVAAANFAVEGSEKDEIAEKLSKDSATITLKALVAK
QAVYAVKDVSKYSGQADIPKALEVIFPNDVAPVSVSPGNVKKVFSKLSKDV
TAVINKPEPATPEENVTITKADVASVYKNGEAVNYTGAADIVGNMMEYEGAT
YKVEKDYTAPEYKDIKVELVDVATILEVTFSELSKQKVVYKKDKVVEGATIALD
QDTTKALVTPGALTDKETANLKVVEVGTKDANNVKNVYEVKTKDVAPNLVK
VVDENKAAATFEDKEVTAOEGRLVYINLDSKDVKEVAVASEDKKAITLTPQEK
GNVKVAATKGFVDTAGNESAAFTKEVYKVEKEEGKDEVARKATVEVAVDSKTF
TVPEDEVKGGQADSASVNYNTLAGAKLPEGLIYVNDGSKSVTLEPETPERS
EYKFTYVYANAKDGYKMGTTNLLVNVVDLKKAPREFSAKTTKDAKEITLTPEEAVY
DATDFVIDLNGVLTATKADERTKASQVYLKATRADVYLAITGYTVAKAEVGVY
LNTADVISINKNKLVAFRPVATR"
286..372
/gene="slpA"
terminator
2767..2801
BASE COUNT 1222 a 447 c 596 g 894 t
ORIGIN
alignment_scores:
Quality: 770.50 Length: 871
Ratio: 1.526 Gaps: 40
Percent Similarity: 57.979 Percent Identity: 32.491
alignment_block:
US-09-754-947-1 x BFH249446 ..
Align seg 1/1 to: BFH249446 from: 1 to: 3159
1 AlagIlystHrPheProAspValProAlaAspHstRgIleAspSe 17
|||||
373 GCAGGAAACATTCCTCCAGACGTCACCTGCTGTGGAGTGCAGAAVA 422
17 rLleAsnTyrLeuValGluLysGlyAlaValLysGlyAsnAspLysGlyM 34
|||||
423 CATTCATTATTTAGTAGCAAAAAAGCAATCGAAGGTAGCCAGACGGA 472
34 ePheGluProGlyLysGluLeuThrArgAlaGluAlaAlaThrMet 50
|||||
473 CATTCGCTCACTGAAGCATCGATCGTCTTCTGCTCGAAATCATG 522
51 AlagIleLeuAsnLeuProIleAspLysAspAlaLysProSerPheAl 57
|||||
523 GCAATTACATTAGTTTGAAGCTTAAGATGCCGCAAAACCATCTTCA 572
67 aAspSerGlnGlyIntPrTyrThrProPheIleAlaValGluLysA 84
|||||
573 ACATGCTCAAGATTCATGGCTGTAATAATATCCACGCTGTTGAACAAG 622
84 lagIValIleLysGlyThrGlyAsnGly...PheGluProAsnGlyLys 99
|||||
623 CTGGTGTAAATTCAGAGTATGGAACACTGTAACCTCAATCAAGCATCA 672
100 lLeAspArgValSerMetAlaSerLeuLeuValGluAlaTyrLysLeuAs 116
|||||
673 ATTAACCGGTCTTATGTCATCAATGATTTAAAGCATACAAATTTAGA 722
116 pHrLysValAsnGlyThrProAlaThrLysPheLysAspLeuGluThrL 133


```

724 AsnVala1GluPheValSerLeuAspThrAsnValVala1GluAsnGlyIynh 740
      :::::::::::::: :::::::::::::: ::::::::::::::
2685 GGAAGAAGTACCTGATCGTAAACGCAACTGCTACCTTCGACGGTAAACA 2734
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2735 TGGT...CGCGTTGGTTATGTAACAGCGTTAAATCAATCATGATACAGTTT 2781
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
757 aLlysAspGlyIysGlu 762
2782 CTTAAGAAAGGAATGAC 2798
seq_name: gb_ba:D86346

seq_documentation_block:
LOCUS      D86346          DNA              07-FEB-1999
DEFINITION Bacillus thuringiensis DNA for crystal protein, complete cds.
ACCESSION  D86346
VERSION    D86346.1  GI:1434919
KEYWORDS   crystal protein; cryTKD.
SOURCE     Bacillus thuringiensis (sub_species:mexicanensis, strain:TKD2-14)
            DNA.
ORGANISM   Bacillus thuringiensis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus; Bacillus cereus group.
REFERENCE  1 (bases 1 to 2842)
AUTHORS   Hashimoto, N.
TITLE      Direct Submission
JOURNAL    Submitted (02-JUL-1996) to the DDBJ/EMBL/GenBank databases. Naoki
            Hashimoto, Hokkaido University, Applied Life Science; N9 W9
            Kita-Ku, Sapporo 060, Japan (E-mail:hashi@eal.hines.hokudai.ac.jp,
            Tel:011-706-2487)
REFERENCE  2 (bases 1 to 2842)
AUTHORS   Hashimoto, N.
TITLE      Bacillus thuringiensis subsp. mexicanensis TKD2-14
JOURNAL    Unpublished (1996)
FEATURES   Location/Qualifiers
            source          1..2842
                        /organism="Bacillus thuringiensis"
                        /strain="TKD2-14"
                        /sub_species="mexicanensis"
                        /db_xref="taxon:1428
                        177..2648
                        /gene="cryTKD"
                        /gene="cryTKD"
                        /codon_start=1
                        /transl_table=1
                        /product="crystal protein"
                        /protein_id="BAI3073.1"
                        /db_xref="GI:1434920"
                        /translation="MAKNTSYKKYIACTMTAAAYAGVSPYAAAGKSPDYQGSNA
                        EYIDYLVAKKATIEGKPGTAPPTAIDIRASAKAMATITLLEKREGAKPSFKDAQDSM
                        AAKTIAAVERAGAVIQGDGTEGFPNNQINIRNSAKTMVKAIKDMQSGQALEKFSDL
                        KDHHGAENKALVALDITNGNGMDEKSVTRAEAKFTAKTDMQSGQALEKVESI
                        KAIKAEIEVLKATVTEETVYKNAKFTLSGQVDVTVADAKIGAKSVLITLEDGR
                        LEKNAALVYVEKLTAEKTEIPALVEIFPFNDVAPTVSVSPDGKVVYSEKLD
                        AAKTVYIVNGQKEGKEVNTPEFNSALNINQNEKPSIIVTGAQTDLADMTMEYBEK
                        ATAYVEDTTTAPREKDKVKEITAEKATLEVTSESLATNGQTVYVKKGEKBEKLT
                        IADAQOKTRAVIEVTGALKEREATDLVEFGYKDAANNVSGSKAVKSKVDVVA
                        PNFKRVADQNKATFTFDEKVEVTAQONKLRVINIDTSQDVYADTVAPADNKGITLL
                        TFPKAGYKIVATKGLVKDAAGNESVAFTVEVKEAEKEDKVAIVKAVAYDKAA
                        TNKTATDEKVEKGVGAESAANVNNYTLAGAKLPEKTLIVLQETKAIIELEPSFTPE
                        KSELVKFTVAVNAKGDVKGCTANLLLEVDTKAPEKSKATIKTVDAKETTPESAV
                        NVDTTDTVDLUNGELTVAKADATAEAKDVLYVLTAPIDVNLATGTVYKAELENN
                        AKTEKTTDAENNLVAFKPTATAR"
BASE COUNT      1118 a          432 c          532 g          760 t
ORIGIN
alignment_scores:      748.50
Quality:
Length:      880

```

Ratio: 1.497 Gaps: 41
Percent Similarity: 56.818 Percent Identity: 31.477

Alignment_block:
US-09-754-947-1 x D86346 ..

Align seg 1/1 to: D86346 from: 1 to: 2842

```

1 AlaGlyLysThrPheProAspValProAlaAspHisTrpGlyIleAspSe 17
   |||||.....|.....|.....|.....|.....|.....|
264 GCAGGAAAAATCATCTCCAGACGTTCAACCTGGTCTTGGATGACAGACTA 313
   |||.....|.....|.....|.....|.....|.....|
17 rIleAsnTrpLeuValGluLysGlyValAlaValIleGlyAsnAspLysGlyM 34
   |||||.....|.....|.....|.....|.....|.....|
314 CATTTGATTATTATGTAGCAAAAGAAAAGCAATCGAAGGTAAGCGACAGCGAA 363
   |||||.....|.....|.....|.....|.....|.....|
34 ePheGluProGlyLysGluLeuThrArgAlaGluAlaIleAlaThrMetMet 50
   ||| ||| : : : : : |||||.....|.....|.....|
364 CATTCGCTCCAACTGAAGCAATTCACCGTGGTCTTCGACGCAAAATTTATG 413
   |||||.....|.....|.....|.....|.....|.....|
51 AlaGluIleLeuAsnLeuProIleAspLysAlaLysProSerPheAl 67
   ||| ||||| : : : : : |||||.....|.....|.....|
414 GCATTATACATTAGCTTAGAGTTAAGAGAGGGTGCAAAACCATCTTTCAA 463
   ||| ||||| : : : : : |||||.....|.....|.....|
67 aaSPserGlnGlyGlnTrpTrpTrpProPheIleAlaValAlaGlyLysSA 84
   |||||.....|.....|.....|.....|.....|.....|
464 AGATGCTCAAGATTCAATGAGGCTGCTTAATATATATGCGAGCTGTGAAGAAG 513
   |||||.....|.....|.....|.....|.....|.....|
84 IaGlyValIleLysGlyThrGlyAsnGly...PheGluProAsnGlyLys 99
   |||||.....|.....|.....|.....|.....|.....|
514 CTGGTGTCATTCAACGGGACGTAAGAACTGGTAGTTAATCCAAACACCA 563
   |||||.....|.....|.....|.....|.....|.....|
100 IleAspArgValSerMetAlaSerLeuLeuValGluAlaIleLysLeuAs 116
   |||||.....|.....|.....|.....|.....|.....|
564 ATTAATGCTGCTCTTATGCGACACATGATCTGAAGCATCAAAATTTAGA 613
   |||||.....|.....|.....|.....|.....|.....|
116 pThrLysValAsnGlyThrProAlaThrLysPheLysAspLeuGluTruL 133
   |||||.....|.....|.....|.....|.....|.....|
614 TCGTAAAGTAAAGCGGACAAATTAGAACTAAATTTCTGTATTTAAAAAGT. 662
   |||||.....|.....|.....|.....|.....|.....|
133 euAsnTrpGlyLysGluLysAlaAsnIleLeuValGluLeuGlyIleSer 149
   |||||.....|.....|.....|.....|.....|.....|
663 . CACTGGGGTGA AAAAGCGCTTAACATCTTAGTTGCTTGGATATCACT 710
   |||||.....|.....|.....|.....|.....|.....|
150 ValGlyThrGlyAspGlnTrpGluProLysLysThrValIleLysAlaG 166
   |||||.....|.....|.....|.....|.....|.....|
711 AATGGTACTGGTAAATGGTGGAGCCAGATAAATCTGTAACCTCGCGAGA 760
   |||||.....|.....|.....|.....|.....|.....|
166 uAlaIleAlaGlnPheIleAlaLysThrAspLysGlnPheGlyThrGluAla. 182
   |||||.....|.....|.....|.....|.....|.....|
761 AGGTGCTAAATTCATTCTTAAGACTGACATCGACTTCGGCAAAAAGGAG 810
   |||||.....|.....|.....|.....|.....|.....|
183 . .AlaLysValGluSerAlaLysAlaValIleThrGluLysValGluVal 198
   |||||.....|.....|.....|.....|.....|.....|
811 AAGCTTAAAGTTGACTCTATTTAAGCAATCATGCTATAAGAAATGCAAGTT 860
   |||||.....|.....|.....|.....|.....|.....|
199 LysPheSerLysAlaValGluLysLeuThrLysGluAspIleLysValThr 215
   |||||.....|.....|.....|.....|.....|.....|
861 AAATCTGCTCTACAGTAACAGAAAGAACTGTGAAGACGCGAAATTTAC 910
   |||||.....|.....|.....|.....|.....|.....|
215 rAsnLysAlaAsnAsnAspLysValLeuValLysGluValIleThrLeuSerG 232
   |||||.....|.....|.....|.....|.....|.....|
911 TCTTTTCCAGGAGCTTATGATACAGACAGTGGTGTGCAAAAACGGGTG 960
   |||||.....|.....|.....|.....|.....|.....|
232 LuAspLysArgSerAla.....ThrValGluLeuTyrSerAsnLeuAla 246
   |||||.....|.....|.....|.....|.....|.....|
961 AAGCCCAAAAATCTGTAATCTTAACTTTAGAGATGTTAAAGAACTTAG 1011
   |||||.....|.....|.....|.....|.....|.....|
247 AlaLysGlnThrTyrThrValAspValAsnLysValGlyLysThrGluVal 263
   |||||.....|.....|.....|.....|.....|.....|
1011 AAAAACAATGCTTATCTTTGTAACGTGTAAGAAAA..... 1044
   |||||.....|.....|.....|.....|.....|.....|
263 IaIleValGlySerLeuGluAlaLysThrIleGluMetAlaAspGlnTrv 280

```



```

224 euVallyscIuVal1ThrLeuSerGIuAspLyArgSerAlaThrValGIu 240
    : : : : : | : : : | | | | | | | : : : |
1326 TTGCATCCGAAGA.....AGTACGATGCTACTTCGCGAGTTACA 1366
    : : : : : | : : : | | | | | | | : : : |
241 LeuTrSerAsnLeuAlaAlaIysGIuThrTrhValAspValAsnLy 257
    : : : : : | : : : | | | | | | | : : : |
1370 CTGGGCGCCAAAGTACTCTCCAAATAAAGACCTTACGTAAAGTG..... 1414
    : : : : : | : : : | | | | | | | : : : |
257 sValGIuLyThrGIuValAlaValGIuSerLeuGIuAlaLyThrIleG 274
    : : : : : | : : : | | | | | | | : : : |
1415 .....AAAATCAATCATCTGTAAACGAAATTCGTATACGAAGGAAAA 1457
    : : : : : | : : : | | | | | | | : : : |
274 LuMetAlaAspGIuThrValValAlaAspGIuPro.....ThrAla 287
    : : : : : | : : : | | | | | | | : : : |
1458 AATTAGACGTAGAANAATCTTACTTGTATGATGATGCGCGCTGTCAACGA 1507
    : : : : : | : : : | | | | | | | : : : |
288 LeuGIuPherThrValLyAspGIuAsnGIuThrGIuValAla..... 301
    : : : : : | : : : | | | | | | | : : : |
1508 ATTGCTTTCAATTAACATGATGAANAAGTAAAGCTGATGTTGAGTACTT 1557
    : : : : : | : : : | | | | | | | : : : |
302 .....SerProGIuGIuYIleGIuPheVal.....T 310
    : : : : : | : : : | | | | | | | : : : |
1558 AACTTAGCAAAACCATGAGCTGAATTTGTACGATTAACTAGACGGTT 1607
    : : : : : | : : : | | | | | | | : : : |
310 hrProAla.....AlaGIuLySLeAsnAlaLyS 319
    : : : : : | : : : | | | | | | | : : : |
1608 CACCAACCAAAACATCTTTGAAGGTGAGAAAGCTACTTCTACTACAGTAA 1657
    : : : : : | : : : | | | | | | | : : : |
320 GYGIuIleThrLeuAlaLySGIuThrSerThrThrValLySAlaValTy 336
    : : : : : | : : : | | | | | | | : : : |
1658 CTAGCTGTGGCATTAAAGCAGGCTGACTCAAGTAGAAGTCAAGTTAC 1707
    : : : : : | : : : | | | | | | | : : : |
336 rLyLyAspGIuLySValValAlaGIuSerLySGIuValLySValSerA 353
    : : : : : | : : : | | | | | | | : : : |
1708 AAAACGCGGCTTTAAAGCTTTCTACACCTGCTATTTACAGGAAAA 1757
    : : : : : | : : : | | | | | | | : : : |
353 laGIuGIuAlaValAlaSerTrSerAsnTrpThValAlaGIuGIu 369
    : : : : : | : : : | | | | | | | : : : |
1758 ACCTGATACACCAAGCTTCTGCATTTAAAAATGTTGTATTGCAATTAG 1807
    : : : : : | : : : | | | | | | | : : : |
370 AsnLySAlaAsp.....PheTrSerLySAs 378
    : : : : : | : : : | | | | | | | : : : |
1808 GCGATAAATGAGTGCTGTAAACATATGCGACCAAGCTTCTGCTGAAGA 1857
    : : : : : | : : : | | | | | | | : : : |
378 rPheLySGIuAsnAsnLyS...ValTrGIuGIuLySAsnAlaTyValG 394
    : : : : : | : : : | | | | | | | : : : |
1858 CTTTGCTTTAAATAGCCAAACCTTACTGTTGTTGTGAANAAGCATCTCTTA 1907
    : : : : : | : : : | | | | | | | : : : |
394 Lh.....ValGIuLeuLySAspGIuPheAsnAlaValTrhTrGIu 407
    : : : : : | : : : | | | | | | | : : : |
1908 ATAAATTAGCTGCTACAAATGTGCTGGAGAAAGATGAAGTAGTTGATCCAGCA 1957
    : : : : : | : : : | | | | | | | : : : |
408 LySValGIuTrGIuSerLeuAsnThrGIuAlaValAlaValAspLySAI 424
    : : : : : | : : : | | | | | | | : : : |
1958 TCATATTAGCATTAATTAACCTTCAAAACAC.....GGTATTATTCTGTAGT 2001D
    : : : : : | : : : | | | | | | | : : : |
424 aThrGIuLySValThrValLeuSerAlaGIuLySAlaProValLySValT 441
    : : : : : | : : : | | | | | | | : : : |
2002 AAATTACTACATTACTGCTGAGGCTGCTGGTGAAGCTCACTTACTATTA 2051
    : : : : : | : : : | | | | | | | : : : |
441 hrValLySAspSerLySGIuLySAlaLeuValSerHisThrValGIuIle 457
    : : : : : | : : : | | | | | | | : : : |
2052 AAGTAGGTAC..... 2062
    : : : : : | : : : | | | | | | | : : : |
458 GluAlaPheAlaGIuLySAlaMetLySAspIleLySLeuGIuLySThrAs 474
    : : : : : | : : : | | | | | | | : : : |
2063 .....GTTCAAAAGCGTTAAATTTAAAGTAAAGCAC 2094
    : : : : : | : : : | | | | | | | : : : |
474 nValAla.....LeuSerThrLySAspValThrAsp..... 484
    : : : : : | : : : | | | | | | | : : : |
2095 TGATTTCGTAAATTAAGTATCATGTTAAAGCTAACCCAGATTAATTACAG 2144

```

485	LeuysValLysAlaProValLeuAlaSerIleu	495
2145	TTGTTCAAAATAAAACAT	TACCTGTTCATCTCGTAACACAGTACCACATAT	2194
496	GlyTysGluPheThrAlaProValThrVal.....	LysValLeuAspLys	510
510	sASpGlyLysGluLeuLysGluGlnLysLeuGlnAlaLysTyrValAsn	527	
2245	AACAGGT...GTACTGTGCAAGGTGGATTTGAGTGTACTAACGACTGCT	2291	
537	rgLLeuLeuVal.....	LeuAsnAlaIaIaGlnGlnAlaGly	539
2292	CTGGTTCATCGGTACAAACAACTGGTGTTCACAGTAATACAGTAGGC	2341	
540	AsnTyrThrValValLeuThrAlaLysSerGlyLysGlnAlaLysAl	556	
2342	GAAAGTACAGTTCACTTC.....CAAAACGTAATGGTGTCTACTTATAGG	2385	
556	arThrLeuAlaLeuGlnLeu...LysAlaProGlyAlaPheSerLysPheG	572	
2386	TTTCATTATATGTGAACGTAAACAGAGGTAAAGTTGCATTATTAACCTTG	2435	
572	Luval.....	ArgGlyLeuAspThrGluLeu	580
2436	AACTTGTATCTAAAGTAGGTCAATATGGCCATACCTGTGATACAAACTT	2485	
581	ASP.....	LysTyrValTrh	585
2486	GACTTAATGTTCACACTACTGTTGATATCATCATATATCTAAGTACACTTC	2535	
585	rgLGlusnGlnLysAsnAlaMetThrValSerValLeuProValAspA	602	
2536	AGATCGGGTATACCTGCTGCTGAAACTTAAAGAGTTATGCAAGTTGAT	2585	
602	IaAsnGlyLeuValLeuLysGlyLysGlnAlaIaIaGlnLeuLysValThr	618	
2586	CTAAAAATCTAGCTGTACCTACGCTCAAAATTTGTTGGAAATAAAAGTTGTT	2635	
619	ThrThrAsnLysGluGlyLysGluValAlaLysAlaThrAspAlaGlnValTrh	635	
2636	GTTACAGGTAAACCTCCACAGTAAGTTGAT.....	2665	
635	rValGlnAsnAsnSerValIleThrValGlyGlnGlyAlaLysAlaGly	652	
2666ATCCACTTAACGAAAAAGGTGCACCTCTGGTA	2699	
652	LuThrTyrLysValTrhValValLeuAspGlyLysLeuLysIleThrThHis	668	
2700	AAGCG...ACAGTGAATATCTTCAAGAGACAAATTGCTATTAAATCTGTA	2746	
669	SerPheLysValValAspThrAlaProThrAlaLys.....	680	
2747	AACTTCAACACAGTTCAACAGAAACTTGTGAGAAAGAAATCAACAT	2796	
681	GlyLeuAlaValGluPheThrSerThrSerLeuLysGluValAlaProA	697	
2797	CGGTACTGTATTAAGAGCTTGAGAGAGTAACTGATGATATCTGTAAG	2846	
697	snAlaAspLeu	700	
2847	GTATTAACTTA	2857	

```
seq_name: gb_ba:AF188935
seq_documentation block:
LOCUS       AF188935                96231 bp            DNA             circular     BCT
DEFINITION  Bacillus anthracis plasmid pXO2, complete sequence.
ACCESSION   AF188935
VERSION     AF188935.1
KEYWORDS    GI:6470151
SOURCE      Bacillus anthracis.
```


ORGANISM	Plasmid <i>Bacillus anthracis</i> Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus/Staphylococcus group; Bacillus; Bacillus cereus group.
REFERENCE	1 (bases 1 to 96331)
AUTHORS	Okimaka,R.T., Cloud,K., Hampton,O., Hill,K.K., Keim,P., Lamke,G., Kumano,S., Mantler,D., Martinez,Y., Svensson,R., Tatum,L.R., Brown,A.E. and Jackson,P.J.
TITLE	Direct Submission
JOURNAL	Submitted (01-NOV-1999) Bioscience Division, Los Alamos National Laboratory, M888, Los Alamos, NM 87545, USA
FEATURES	Location/Qualifiers
source	1. 96231
CDS	<pre>/organism="Bacillus anthracis" /plasmid="px02" /db_xref="taxon:1392" complement(19..477) /codon_start=1 /transl_table=11 /evidence=not_experimental /product="px02-01" /protein_id="AAFI3606.1" /db_xref="GI:6470152" /translation="MREGGDDVSLKKEVKDLKQSYNTMKQOMDLIKLNMELIAQ SGSNLSLEKLEQEVKSEIASVYTKSEGGFPPTVYKREPIEENVOEHQRIQREYKTTSF VAKPLMITDDVQIYMKKGKPHQVPTRLGTQIETEMEQIPEHRRKELE" complement(482..1018) /codon_start=1 /transl_table=11 /evidence=not_experimental /product="px02-02" /protein_id="AAFI3607.1" /db_xref="GI:6470153" /translation="MGKSPSIIILTSQYTIIPNKSFTTYVDYCMROKALLKDGHTLNT ELTELRLQNALIDSEPMERGSPHSNOKVYSEKEAKVLLTSPNSFPTTDFEKYV SYMGKRYAESKKKMTKTEEEVSLNKHIEDLSPMEKGEKKGKABEILSGVRSK DKNVITSTDLNFKDRIK" complement(1037..1633) /codon_start=1 /transl_table=11 /evidence=not_experimental /product="px02-03" /protein_id="AAFI3608.1" /db_xref="GI:6470154" /translation="MTYKLFQFRNQIHVDIVDFORQYPRPADVLRHAIYLLD EKMEREGGDDVSLKKEVKDLKQSYNTMKQOMDLIKLNMELIAQSGSNLSLEKLEQ EVKSEIASVYTKSEGGFPPTVYKREPIEENVOEHQRIQREYKTTSFVKKPLMITDDV QIYMKKGKPHQVPTRLGTQIETEMEQIPEHRRKELE" complement(1645..1932) /codon_start=1 /transl_table=11 /evidence=not_experimental /product="px02-04" /protein_id="AAFI3609.1" /db_xref="GI:6470155" /translation="MEIHVRNDVPHLKEIDRKCRIGKGLGRYRYRWETINMFEQH FNOEISRNKEDKDEAVSNVSTILDKQSKRLQRYIDATHELVAAMTKLNEE" complement(2240..3088) /codon_start=1 /transl_table=11 /evidence=not_experimental /product="px02-05" /protein_id="AAFI3610.1" /db_xref="GI:6470156" /translation="MSIKFPMNKEGRKPANTKRAYLFTLGSFVTMFVLCISPVPSG ATYKFEEMKIGEQSLSTYKIAVAKENYNDQTLRIDYELRADNSOILSMKKYK ENKYIKQDNVNTKYVRASDNTIYVISENVPEEGVSVYKPEYIHPELDQVDVL KRSKMYVLENKILNRLKRSKSDPYREVIAPSOALRKIEKKMKDSAMQOL KIKNEQJLKEMEQTEGKRYKTKNTINSNESTINNHQEKIDVLKEDIKKKKTIQLD EKKKTI" complement(3085..3360) /codon_start=1 /transl_table=11 /evidence=not_experimental</pre>
CDS	<pre>/product="px02-06" /protein_id="AAFI3611.1" /db_xref="GI:6470157" /translation="MLSELKPNHDYAKGKYLILSRKKGVRRQKFIETPTWFDY NCEGEVEMIVREIOPVSGKEKRYNCKLEINHQAQVSVYVNGERMK" complement(3449..4057) /codon_start=1 /transl_table=11 /evidence=not_experimental /product="px02-07" /protein_id="AAFI3612.1" /db_xref="GI:6470158" /translation="MNNLSKITSGALVLSLSPNGIMAYNTTEPDKDKKRYVYSEK EKQIDGLKYLEQKCKINEGAGENKSTLIDQNTQKREVAQFHAIVYISVON KEERNNLKILTDKRVKVIDIAPNTDLDGPNKSHVKNKAIITYINSGDVSKCTALD IEYTTGLENKQTTINSYVKITLERQGEIEIKVEYNPYVKR" complement(4083..5225) /codon_start=1 /transl_table=11 /evidence=not_experimental /product="px02-08" /protein_id="AAFI3613.1" /db_xref="GI:6470159" /translation="MAAQETAIIDKYKKYKRIKWIVRLIGSTGVVIAAITLLIIVS MAIFGQSSTGTGPGISGTATYKNIIPPEYMKQANVBOGCAAGQPELIVYLAIIIM VESNGISEKLPDIMOSSESQGWAMNTISPKOSIYGVNHLGAFDDAKMLGINDLLA IVQYTNFGRNVVHMLAANNKTHSIQIADYSLTVVAPAGNNGNTTIGVSOVAAVYN GGYRIYINGNFFYAEWKQYLSFDGAGTSGQIPGSEFEKVMQDEVLKYNGNPYVMG GSSSGGFDPCSGLTWYKRTAGITIPISATQIDFTVEVDPAQDGLVFPFGYTG PMHVSHVGIYDIAFMTPDSCGSGVGHQPTSSYMOQHVAGIRVVR" complement(5227..7158) /codon_start=1 /transl_table=11 /evidence=not_experimental /product="px02-09" /protein_id="AAFI3614.1" /db_xref="GI:6470160" /translation="MKLGKKEQOKYRTBGYDIDFTISRVQPGGKIFNEKYIAGDCYV ACLIHVSLAEDIPLMLTALMNNKDTISSVDVATNKEEYVNDINKSTIELKDRMTSE RSTDPDAMHMLONLDFARSITQOGEIYVAVKRLIYIPDLQLEKIDIDKEI ACQNKYAGVYFROKEMOTLPASVDQIEYGVSGYPLPKSNGIFGIPFHQDKD PBGIVGORTSGTGAFILDPFSTGRTSPGTFPGMGKATLSTLKOLEGVLVADCE HGFDPKARYTYVVOQGGKIIDLSGVDSQSEVETGNTMPLSLPKKLTNTEGAVDCE SETQHSKYVTNMRILNTEFDVTVIOEERKHLAPFTIENGLPQKSDRPYVVTGTP NMYPLIEDDYKYLSNLKLEAGATPOHNRDLIAIKLOVEDMTIYVGDMEFGHTLNFE NQOIVYFEDGYSKYDSYNGCOLFTPALILMSHALNCRQKUYLREEKNLSIEDVKY FMVFLDECHNVINSQNEFANRYVLEFERMRKFPAGIFPAQOSQOMLPPPGSTYDV TMKAIPELTQYKIFLNMNSVLETLKSVLGESELTSEERILPELKRGEAIVQVSTET YVMEDPDEKOLERFKGQ" complement(7178..7846) /codon_start=1 /transl_table=11 /evidence=not_experimental /product="px02-10" /protein_id="AAFI3615.1" /db_xref="GI:6470161" /translation="MKVATAEAKKYNRKQIIESEKNGKOSTISTIPASFKNFTASIV DIAPYKRLQOKTGILVDHRNNLOVYLKVTTLILSNQODLKRPMQNLSTLCRVYHEP FKILSTYSTETTEQOYVYKRMALRFQGMSEVSEKKEEHLAYORYSLAENLNKLV WVENKLELAFPIVYVGKNETELIKVKNKRYGGRQFNLQNMKAKEVERKLLFKLQNM NSEM" complement(7910..8224) /codon_start=1 /transl_table=11 /evidence=not_experimental /product="px02-11" /protein_id="AAFI3616.1" /db_xref="GI:6470162" /translation="MEMMRNPKNTKQEIKLAFYIITDGAIIALMLVASYMPKVPVG GGRIMFYVLFSTFGFLICIKPHNSPTNNIFVYIIDLMLKMDKNYHPFVNTISSETK RK" complement(8227..8430) /codon_start=1</pre>

[illegible]

```

96 ProAsnGlyLysIleAspArgValSerMetAlaSerLeuLeuValGluAl 112
114 CCAAGTACACATCATCTGCCCAAGAAATGGCAAAATGGTAGTAGAAGC 1213
112 ATTrLysLeuAspThrLysValAsnGlyThrProAlaThrLysPheLysA 129
1214 TTATACCTTGAGCTA...GTTGAAGGTTAAAGAC...CTACCATTCACCTG 1257
129 sPLeuGluThrLeuAsnTrpGlyLysGluLysAlaAsnIleLeuValGlu 145
1258 ATGTTAGCGGCTCT...TGGAGTACTGACTACATCAACATCCTGCTCTCT 1304
146 LeuGlyIleSerValGly...ThrGlyAspGlnTrpGluProLysLysTh 161
1305 AACGGCGTAGCGCTGGTATGACACAGCACTACATTGGCAGCAGCGCGGA 1354
161 rValThrLysAlaGluAlaIleAlaGlnPheIleAlaLysThrAspLysGlnP 178
1355 AGTACTACGCTGCTCAACACAGCAGCATTTATCCACCCTGCAGAG... 1397
178 heGlyThrGluAlaAlaLysValGlu... 186
1398 ...GTTCAAGAAAGAGCGGTATCGAAAGTTCCGGGCTAGTTCCTACAGTT 1442
187 ...SerAlaLysAlaVal... 191
1443 TCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1492
192 ...ThrThrGlnLysVal... 196
1493 CGATTACTCTCTGAGTAAGTACTCTGACACAGCTGCTCTTACTAGTAAAG 1542
197 ... 204
1543 CTACAGAGTACTCTTTCAGAGATTACAGATGTTAGTATAGTACAGAGCTG 1592
205 GluLysLeuThrLysGluAspIleLysValThrAsnLysAlaAsnAsp 221
1593 ACTGGGTTTATGAGAAATAGCTGTACATCTGTGAATGCTGTGAATGG 1642
221 pLysValLeuValLysGluValThrLeuSerGluAspLysArgSerAlaT 238
1643 AAACAATTAAGTGTATTTACACACACACAGCTGATGCTAGTACAGTAT 1692
238 hrValGluLeuTrpSerAsnLeuAlaLysGlnThrTrpThrValAsp 254
1693 TAAGTCTCTCTACAGCTTCTACTAGCACTGTCAAAACGTGTACTACT 1742
255 ValAsnLysValGly...LysThrGluValAlaValGlySe 267
1743 TTAAACGACATCTACTGACACACATAGTAAATGATGTTGTTTAAAGTGT 1792
267 rLeuGluAlaLysThrIleGluMetAlaAspGlnThrValAlaAlaAspG 284
1793 TGCAGAACGAGCTGTACTAAGTCTGATGGGAAAAACACTTACAGTAAG 1842
284 LuProThrAlaLeuGlnPhe...ThrValLysAsp 294
1843 CGCCTAGCTGTCAGTCTTCAAGGCAACTATGATGTAAACAGTTAAGAT 1892
295 GluAsnGlyThrGluValValSerProGluLysIleGluPheValThr... 310
1893 GCAAAGACGGCTCAAAACACAAATTCAGTGTAAACAAATTTCTGCT 1942
311 ...ProAla 313
1943 AGAGAGACACTACAGCTCAATTGTAACACTCAGTGTATTATCTACTGCA 1992
313 laGluLysIle...AsnAlaLysGly 320
1993 CTGATTAATTTGAAGTACTTATACAGAGCTATTGACTCATTAACGGGT 2042
321 GluIle... 322

```

```

2043 GAAGTATTACGTATTAAAGGTCACACCTGTGCTAGTGGGTTGAGCTTT 2092
323 ...ThrLeuAla 326
2093 AACGAGACCAACAATAATTATACATTGCTGCTGCTTCATCTGTGCTA 2142
326 ysGlyThrSerThrThr... 331
2143 CTGATTAATTAACCAACAATTATATCGCTGATTTAGTATGAGCGGAGGT 2192
331 ... 331
2193 AACTTTGTACTCCAAAGTACTACACAGTTCCTGTTACTCAAGATACAC 2242
332 ...ValLys 334
2243 TGCTTTAGCTGTGCTCATCTTAGAGCAAGTACTAATCAAAAAGTAGCTT 2292
334 laValThrLysLysAspGlyLysValAlaIleGluSerLysGluValLys 350
2293 TAACATTCATTAAGAA...TTAACTCAGCATCTTAAGACGGCATTG 2336
351 ValSerAlaGluGlyAlaAlaVal...AlaSerIleSe 362
2337 GAAAGTGTACAGAGAGTAGTGTGTTACTGCTCCAAATGGAAGACTACATC 2386
362 rAspTrpThrValAlaGluGlnAsnLysAlaAspPheThrSerLys... 377
2387 AAATTTCAGAGTGTCTCAAAATCTTACAGTAGACATATCGTAAACTT 2436
378 ...AspPheLysGlnAsnAsnLysValTrpGluGlyAspAsnAlaTrpAl 393
2437 ATGATATTACATTAAAGTGATGCTACTATACAAAGTAACCTGCAAGT 2486
394 ...GluValGluLeuLysAspGlnPheAsnAlaValThrThrGly 408
2487 TTTGGTATCAGCTTAATTAAAGATGCAATTACGAGCGTACTAAATVAA 2536
408 sValGluTrpGluSerLeuAsnThrGluValAlaValAlaAspLysAlaT 425
2537 A...AATGACCTTTATTCAAATTCATTTCTTAAATAAAGATA 2577
425 hrGlyLysValThrValLeuSerAlaGlyLysAlaProValLysValThr 441
2578 CTGATCAGCACTAGTCTCTCTGCGCTGAGCTAGTAAACAGCAACA 2627
442 ValLysAspSerLysGlyLysAlaLeuValSerHisThrValGluIleG 458
2628 ATC...GAAGTAACCTTAAGTAGAGGGTGTACTATTAC 2662
458 uAlaPheAlaGlnLysAlaMetLysAspIleLysLeuGluLysTrpAsnV 475
2663 AAACCCAGACACAGTAAGTAATTAAGTAAGATGCTGCTGAGCAACGGAT 2712
475 aAlaLeuSerThrLysAspValThrAspLeuLysValLysAlaProVal 491
2713 TATCTGTGCTTTAAAGGCTGTACTGAT... 2741
492 LeuAspGlnTrpGlyLysGluPheThrAlaProValThrValLysVal 508
2742 ...AACGTATTAGTAGT 2755
508 uAsp...LysAspGlyLysGluLeuLys...GluLysL 520
2756 CTCTTATAGCAGCTGAGCTGAACTAGCAGGGCGCTTATCAAGTACGTT 2805
520 euGluAla...LysTrpVal 525
2806 TAGAGCGAGCGAGCATTAACATTTAAATGTAATGTAATTAATGAGGTT 2855
526 AsnArgGluLeuVal...LeuAsnAl 533

```

2856 ATGACCTTCAGTTCTGTCTGCTCAACTCCAGCTGCTCTTAAATGT 2905
533 aAlaGlyInGluAlaGlyAsnThrValValLeuThrAlaLysSerG 550
2906 TGCTGTGCTAATTCTGGCAGCTAACGATTATTACTGTAAGTGCACCACTG 2955
550 lGlyLysGluAlaLysAlaThrLeuAlaLeuGluLeuLysAlaProGly 566
2956 GACAAACATTTACTGCTGCTACCTA..... 2981
567 AlaphSerLysPheGluValArgGlyLeuAspThrGluLeuAspLysTy 583
2982 AACCTTAATAATTAAATCGATGCGAGCTGTAACTCAACTCGCA 3031
583 rValThrGluGluAsnGluLysAsnAlaMetThrValSerValLeuProV 600
3032 TATTACTTAATTTCTACTCGTATATTATACAGATATCTTACCATG 3081
600 aAlaPala.....AsnGlyLeu 605
3082 AGGACTCGGTTAAATTTCTGTAATGCTTATTACTACTAATAGTTTA 3131
606 ValLeuLysGlyAlaGluAlaAlaGluLeuLysValThrThrAsnLy 622
3132 GCATTA.....GA 3139
622 sGluGlyLysGluValAlaPalaThrAspAlaGluAlaThrValGluAsn 639
3140 ATCCGGAAGAGCTTGTGCTACTGCTACTGCTACTGCTACTGTTACTG 3189
639 snSerValIleThrValGlyGlnGlyAlaLysAlaGlyLuhThrLys 655
3190 AT..... 3191
656 ValThrValValLeuAspGlyLysLeuIleThrThrHisSerPheLysVa 672
3191 3191
672 lValAspThrAlaProThrAlaLysGlyLeuAlaValGluPheThrSerT 689
3192ACTGCTCCAACTTTAAACAGGTGCACAACTTA..... 3221
689 hrSerLeuLysGluValAlaProAsnAlaAspLeuLysAlaAlaLeuLeu 705
3222GTTAGTGCACAAACGTAAT 3239
706 AsnIleLeuSerValAspGlyValProAlaThrThrAlaLysAlaThrAl 722
3240 AAGTTCACCTTCGATGAGAAATTAGTCTTACTCTTACTACTAGATGCTGC 3289
722 aSerAsnValGluPheValSerAlaAspThrAsnValValAlaGluAsnG 739
3290 TGAATTAATTGATGATTCATTCATTAACAAAGCTACAGTTGCTCAATG 3339
739 lYThrValGlyAlaLysGlyAlaThrSerIleTyValLysAsnLeuThr 755
3340 GTGGTTTT...GCATCCGAGCAACTCT.....GGTGAGAGTACT 3377
756 ValValLys.....AspGlyLys 761
3378 GTAGTATCTTCAGTATGCTGTA 3401
seq_name: gb_ba:AF055578
seq_documentation_block: 3680 bp DNA BCT 08-MAR-2000
LOCUS AF055578 Bacillus stearothermophilus surface layer protein precursor (sbpc)
DEFINITION gene, complete cds.
ACCESSION AF055578
VERSION AF055578.2 GI:4581955
KEYWORDS Bacillus stearothermophilus.
ORGANISM Bacillus stearothermophilus.

REFERENCE
AUTHORS Jarosch M., Egelseer E.M., Mattanovich D., Sleytr U.B. and Sara M.
TITLE S-layer gene sbpc of Bacillus stearothermophilus ATCC 12980: molecular characterization and heterologous expression in Escherichia coli
JOURNAL Microbiology 146 (Pt 2), 273-281 (2000)
MEDLINE 20170659
PUBMED 10708365
REFERENCE
AUTHORS Jarosch M., Egelseer E.M., Mattanovich D., Sleytr U.B. and Sara M.
TITLE Direct Submission
JOURNAL Submitted (25-MAR-1998) Zentrum fuer Ultrastrukturforchung, Universitaet fuer Bodenkultur, Gregor Mendelstrasse 33, A-1180 Vienna, Austria
REFERENCE
AUTHORS Jarosch M., Egelseer E.M., Mattanovich D., Sleytr U.B. and Sara M.
TITLE Direct Submission
JOURNAL Submitted (13-APR-1999) Zentrum fuer Ultrastrukturforchung, Universitaet fuer Bodenkultur, Gregor Mendelstrasse 33, A-1180 Vienna, Austria
REMARK
COMMENT Sequence update by submitter
FEATURES
On Apr 13, 1999 this sequence version replaced gi:3025825.
source
Location/Qualifiers
1..3680
/organism="Bacillus stearothermophilus"
/strain="ATCC12980"
/db_xref="taxon:1422"
/db_xref="ATCC:12980"
51..3643
/gene="sbpc"
51..3643
/gene="sbpc"
84
/gene="sbpc"
/note="alternative transcription start site"
255..261
/gene="sbpc"
273..362
/gene="sbpc"
273..3572
/gene="sbpc"
/note="S-layer protein; Sbsc"
/codon_start=1
/transl_table=11
/product="surface layer protein precursor"
/protein_id="AAC12757.1"
/db_xref="GI:3025825"
/translation="MDKKAVKLATSAVAASFVAANPASAQATDVATVSOAKAO MKDAVYVTSHTVETEGQPDIKDYVAAYNKAQAYANAVALVKKAGAKKADVLADQ AIETEVFKANPKSGEARVATYIDAYVYAKLRQELAAADAKDKAEELYHKI SYLKRPTVLIDVYCGSTRRELIRSTFKAAQALRBLIDIVAMKARAPDAVAG NDKARAKDOVNOVYKVDARFAELOKAQAQAKAEALAPKVESVAIDSTSEK VFTTKPVDKATAPKNSITLKTERTLKYKSVSSGSLTATVTLIDPLVDKSTYV VTGDKDTAKKEFEETNEFTYKPKVPASITFENPKLPESAVDLTKYTVYKRAAGV IKSGFELETSSEKLTKQKEINTGKSVLVNATVGTNVTGNVLIAEDEKAAEVS ELRLTKDNKEVTLVANGNAFDRDQNOISGTLTLAKFDDQGNELTGKVAQDTYF ESINPELVVAPDGSVTP IYVGTALVYKGEVTKTIPVYKANPYLETLAVDSTVS VAKGOKATFEVTLKDOYGKNEFTGNVNTSDKTEFATVSNSGIGSEVYVTVGAE GSTTITIKSTKEKYKPVNVYVAGPANYGIKYLDQCKIDKSATPESANNDOLKYVA VDAANGNLVGDITNDVITTSATDITNGVIVNASKSTANGDTIYITDNGSKVKKEKIT VKLGYTLGTVADVEVITDLKATVYVYKKADELIDADNDALAKLILANDKIDQNGN PMVDSATPMTNEKLQALKSGLVLSGIIVSDSVGSVNDLADDAISGLAVAKAGT VTLTIVNEDSKLAPIAITVYKAPAAQDDGVYGLDVPVGTGVTGTFATDRIKSG HKLYAVNDSDAVAPAVAGTTRNSGKFNANETVGTTEVAARAQGIITVIEVDSSDRVG YKFTVYENADLSVADKGTFTATVPYTPGSGQVYTGKTLILAVSPLANCHKLYAAAGSS AAAPVKGIVATVDTVYKTTVGTETVSGVEVDQDGOHSITIEVDENGKVVYKIDYTI TGIOIGTKSAS"
663..3569
/gene="sbpc"
/product="surface layer protein"
3606..3643
terminator

BASE COUNT 1265 a 694 c 766 g 955 t
ORIGIN

alignment_scores:
Quality: 445.00 Length: 908
Ratio: 0.939 Gaps: 41
Percent Similarity: 52.203 Percent Identity: 24.559

alignment_block:
US-09-754-947-1 x AF055578 ..

Align seg 1/1 to: AF055578 from: 1 to: 3680

```

44 AlaGluAlaAlaThrMetMetAlaGlnLeuAsn..... 55
   ::::::::::::::::::::
354 TCCCAAGCGCAACGCGAGCTGGCGAGCGTCTCAGCCAAAGCCAAAGCCGA 403
   ::::::::::::::::::::
55 ..... 55
404 AATGAAGAAGCATACTATACGTACAGCCATACGGTAACGAAAGCGGCC 453
   ::::::::::::::::::::
56 ..LeuProIleAspLysAspAlaLysProSerPheAlaIleAspSerGlnGly 71
   ::::::::::::::::::::
454 AATTCGCCGACATTAAAGATGTAATGCCCGCTACAAACAAACAAA... 500
   ::::::::::::::::::::
72 GlnTPryrThrProPheIleAlaValGluLysAlaGlyValIleLys 88
   ::::::::::::::::::::
501 CAAGCGTATGCAAAAGCGGTTCTGCTGCTGCAACAAAGCGGCGGCGGA 550
   ::::::::::::::::::::
88 sGlyThrGly.....AsnGlyP 94
   ::::::::::::::::::::
551 AAAAGACGCGTATTTGGCTGATTGCCAGCTATATACGAACGTAATGTA 600
   ::::::::::::::::::::
94 heGluProAsnGlyLysIleAspArgValSerMetAlaSerLeuLeuVal 110
   ::::::::::::::::::::
601 TCAAAAGCCAAACCGAAATCGGGCGAAGCGCGTGGCTACG...TACATC 647
   ::::::::::::::::::::
111 GluAlaTyrLysLeuAspThrLysValAsnGly.....Th 122
   ::::::::::::::::::::
648 GACGCTTCAACTATGCGACAAATTTAGACAAATGCGCCAGAACTCA 697
   ::::::::::::::::::::
122 rProAlaThrLysPheLysAspLeuGluThrLeuAsn..... 134
   ::::::::::::::::::::
698 AGCAGCGCTGCATGCGAAAGATGTGAAAAAGACGAAGAGCTTACACA 747
   ::::::::::::::::::::
135 .....Trp 135
748 AATTTCTGACAGCTCAAAAGCGCGACGGTCATTCTGACCGCGTTCAC 797
   ::::::::::::::::::::
136 GlyLysGlu..... 138
   ::::::::::::::::::::
798 GGCMAATCGACGCGCGAATCCTCCGTTGAGGTCAAAAGCCGATGTC 847
   ::::::::::::::::::::
139 .....LysAlaAsnIleLeuValGluLeuGlyIleSerValGlyThrG 153
   ::::::::::::::::::::
848 AGCATTTGCGCAACAGATTGATCTACGACATCACCGTTCATGAAAGCGC 897
   ::::::::::::::::::::
153 LysAspGlnTrpGluProLysLys.....ThrValThrLysAlaGluAla 167
   ::::::::::::::::::::
898 GTGAAGCCCAAGACCGCTGTCAAAAGCGCGCAACTTGACAAAGCAAGCC 947
   ::::::::::::::::::::
168 Ala.....GlnPheIleAlaLysThrAspLysGlnPheG 179
   ::::::::::::::::::::
948 GCGCTTGACCAAGTCAACCAATACGTATCGAAAGTAACCGATGCGTTCAA 997
   ::::::::::::::::::::
179 yThrGlu..... 181
   ::::::::::::::::::::
998 AGCGAAGCTGCAAAAAGCAGACACAAGACCGCAAGCGCGCTATGAAACAG 1047
   ::::::::::::::::::::
182 .....AlaAlaLysValGluSerAlaLysAlaValThrThrGlnLysVal 196
   ::::::::::::::::::::

```

```

1048 CATTGACGCCGAAGTTGAGAGTGAAGTGAAGTGCATTCCTACAAAGCTTC 1097
   ::::::::::::::::::::
197 GluValLysPheSerLysAlaValGluLysLeuThr.....LysGlnAs 211
   ::::::::::::::::::::
1098 AAGGTACATTACTTAAGCTGTAGCAAAAGCAACAAACAAATTCCTAAAAA 1147
   ::::::::::::::::::::
211 PileLysValThrAsnLysAlaAsnAspLysValLeuValLysGluVal 228
   ::::::::::::::::::::
1148 CTTCCTATTACTTTAAAGAGTACGAACTAACTTATCTCTAAGTCTCG 1197
   ::::::::::::::::::::
228 aThrLeuSerLysAspLysArgSerAlaThrValGluLeuTyrSerAsn 244
   ::::::::::::::::::::
1198 TTGAAGTATCTGAGAGTGATTAAGTCTACTGTAACCTTTATGATACA 1247
   ::::::::::::::::::::
245 LeuAlaAlaLysGlnThrTyrThrValAspValAsnLysValGlyLysTh 261
   ::::::::::::::::::::
1248 TTAGTAGACGCTAAACCTTACACTGTGTGAACATCT.....GGTTTAAA 1291
   ::::::::::::::::::::
261 rGluValAlaValGlySerLeuGluAlaLysThrIleGluMetAlaAspG 278
   ::::::::::::::::::::
1292 AGATACAGCTGTGAAGAATTTGAAACTAGCACAAACGAATTCACATTACA 1341
   ::::::::::::::::::::
278 lntThrValValAlaAspGluProThrAlaLeuGlnPhe..... 290
   ::::::::::::::::::::
1342 ACAAGCCAGTA.....CCAGCTTCATTATCAATTCAACTCAATAG 1382
   ::::::::::::::::::::
291 .....ThrValLysAs 294
   ::::::::::::::::::::
1383 TTACCTGAAGATAGTGCAGTATTCATTTGACTTAATACGTACTGTAAAGA 1432
   ::::::::::::::::::::
294 pGluAsnGlyThrGluValValSerProGluGlyIleGluPheValThrP 311
   ::::::::::::::::::::
1433 TGCTCTGTGTACGTAAATTAATCT.....GGCTTCAGATTAGAATCA 1476
   ::::::::::::::::::::
311 roAlaIleGluLys.....IleAsnAlaLysGlyGlu 321
   ::::::::::::::::::::
1477 CTTCAGTGAATAAATTAACCTCAAGTAAATTCATATCACTGCTGTAAAG 1526
   ::::::::::::::::::::
322 lIeThrLeu.....AlaLysGlyThrSer...ThrThrVa 332
   ::::::::::::::::::::
1527 AAGTCAAGTTATTTGTAAAGCTACTGTAAAGGCCAAATGTTACAACTGG 1576
   ::::::::::::::::::::
332 lLysAlaValTyrLysLysAspGlyLysValAlaLacGluSerLysGluV 349
   ::::::::::::::::::::
1577 TAACCTAATCCTTCGAGTGAAGATGAAGAAAGCTCAAGAAAGTTAGTAAT 1626
   ::::::::::::::::::::
349 aLysValSerAlaGluGluLysAlaValAlaSerIleSerAsnThrPthr 365
   ::::::::::::::::::::
1627 TAAATTTAAGTAAAGCAACAAAGAGTTGTAACTCTT..... 1664
   ::::::::::::::::::::
366 ValAlaGluGlnAsnLysAlaAspPheThrSerLysAspPheLysGlnAs 382
   ::::::::::::::::::::
1665 TACGCAAAATGTAACGCAATTTGAT.....AAAGATGG 1696
   ::::::::::::::::::::
382 nasLysValTyrGluGlyAspAsnAlaTyrValGlnValGluLeuLysA 399
   ::::::::::::::::::::
1697 TAACCAAAATCTCTAGCGGT...ACTTTAACATTAAACAGCTAAGTTCAAG 1743
   ::::::::::::::::::::
399 sPglnPheAsnAlaValThrThrGlyLysVal.....Glu 410
   ::::::::::::::::::::
1744 ACCAATATGTAATGAATTAATTAAGTGAAGTACGTAAGTACTATATCA 1793
   ::::::::::::::::::::
411 TyrGluSerLeuAsnThrGluValAlaValAlaLysLysAlaThrGlyLys 427
   ::::::::::::::::::::
1794 TTTGAATCTTAATTCCTGAAGTGTAGTACTT...GCTCTGAATGTAG 1840
   ::::::::::::::::::::
427 sValThrValLeuSerAlaGlyLysAlaProValLysValThrValLysA 444
   ::::::::::::::::::::
1841 CGTAACCTCAATCTTCACGCTACTGACACTGTGAAGTTAAATATGTTG 1890
   ::::::::::::::::::::
444 sPserLysGlyLysAlaLeuValSerHisThrValGluIleGluAlaPhe 460
   ::::::::::::::::::::
1891 AG.....GTACTAAGACAAATCCCTGTGACTGTAAA 1922
   ::::::::::::::::::::

```


alignment_block:

US-09-754-947-1 x A93836 ..

Align seg 1/1 to: A93836 from: 1 to: 2766

```

4 ThrpProAspValProAlaAspHisTrpGlyIleAspSerIleAsnTy 20
   : : : : : : : : : : : : : : : : : : : : : : : : : :
97 ACCCTCACAGATGTTGCCGCCGAATAT.....AAGATGCATGATTT 140

20 rLeuValGluLysGlyAlaValLysGlyAsnAspLysGlyMetPheGluP 37
   : : : : : : : : : : : : : : : : : : : : : : : : : :
141 CTTAGATCAACTGCTGTCACAAAGGTAAGTAACAGAACTAATTCGGCG 190

37 rGGLyLysGluLeuThrArgAlaGluAlaIleThrMetLeuIle 53
   : : : : : : : : : : : : : : : : : : : : : : : : : :
191 TTTACGATCAAAATCACTGCTAGATGCGCGCATTAATCTTCAAGACTA 240

54 LeuAsnLeuProIleAspLysAspAlaLysProSerPheAlaAspSer 70
   : : : : : : : : : : : : : : : : : : : : : : : : : :
241 TTTAAACTAGACGTTGACAAAGCAAAAGACGACGCTTCACGATGTGCC 290

70 nGlyInTrpTyThrProPheIleAlaValGluLysAlaGlyValI 87
   : : : : : : : : : : : : : : : : : : : : : : : : : :
291 AAAAGAC...CGTGCAAAATAGTCACGCGCTTGTAAGAGCTGGCGTAT 337

87 IeLysGlyThrGlyAsnGly...PheGluProAsnGlyLysIleAspArg 102
   : : : : : : : : : : : : : : : : : : : : : : : : : :
338 TTAACGGGTAAAGCACCTGGCAATTTGTGTACGACCCATTAACTGCTC 387

103 ValSerMetAlaSerLeuLeuValGluAlaIleTyLysLeuAsp..... 116
   : : : : : : : : : : : : : : : : : : : : : : : : : :
388 GTTGAATATGCAAAAATCATCGCGAACCCTTACAAATTTAAAGCTGACGA 437

117 .....ThrLysValAsnGlyThrProAlaThrLysPheL 128
   : : : : : : : : : : : : : : : : : : : : : : : : : :
438 TGTAAACTTCATTCATGATGATGAACGATACGCGACCATACGTAA 487

128 yAspLeuGluThrLeuAsnTrpGlyLysGluLysAlaAsnIleLeuVal 144
   : : : : : : : : : : : : : : : : : : : : : : : : : :
488 AAGCCCTT.....TATAATACGAAGTAACCAAAAGGTTAAAA 525

145 GluLeuGlyIleSerValGlyThrGlyAspGlnTrpGluProLysThr 161
   : : : : : : : : : : : : : : : : : : : : : : : : : :
526 CACCAACAAGCTTCGCTGCATAC.....AAAAA 554

161 rValThrLysAlaGluAlaValGlnPheIleAlaLysThrAspLysGlnP 178
   : : : : : : : : : : : : : : : : : : : : : : : : : :
555 CATCACTCTGCGTACTTTCGCAATTTGTATATAGCGGTGAATATT. 603

178 heGlyThrClnAlaIleLysValGluSerAlaLysAlaValThrGln 194
   : : : : : : : : : : : : : : : : : : : : : : : : : :
604 .....AATCGAGTGCACAAATAGTTGAAGTAACGCGTTAATTGCACT 648

195 LysValGluValLysPheSerLysAlaValGluLysLeuThrLysGluAs 211
   : : : : : : : : : : : : : : : : : : : : : : : : : :
649 ACAGTGAAGTAACATTCATACGCAATTT.....GCTGA 683

211 rIleLysValThrAsnLysAlaAsnAspLysValLeuValLysGluV 228
   : : : : : : : : : : : : : : : : : : : : : : : : : :
684 TGTGATTTTCACAAAATTTTGTATGATACGATTAACTGTTACTAAG 733

228 aThrLeuSerGluAspLysArgSerAlaThrValGluLeuTySerAsn 244
   : : : : : : : : : : : : : : : : : : : : : : : : : :
734 CAACCTTTCTCGTGATAAAAATCCGTAGAGTGTGTGAATTAACCG 783

245 LeuAlaIleLysGlnThrTyThrValAspValAsnLysValGly.... 259
   : : : : : : : : : : : : : : : : : : : : : : : : : :
784 TTTACTCTTAATCAGGAATTAATACACGACGACGCAATTAAAAATTT 833

260...LysThrGluValAla.....ValGlySerLeu... 268
   : : : : : : : : : : : : : : : : : : : : : : : : : :
834 AAAAGGCCAGACCGCTAAGAAATTAACGTGTAAGTTTGTGGTCTGTC 883

```

```

269...GluAlaLysThrIleGluMetAlaAspGlnThrValAlaIleAspGln 284
   : : : : : : : : : : : : : : : : : : : : : : : : : :
884 AAGATGCGGTAACTGTTGCCTAAATTAATATGTTCCGTTAACTGAGAG 933

285 ProThrAlaLeuGlnPheThrValLysAspGluAsnGlyThrGluValVa 301
   : : : : : : : : : : : : : : : : : : : : : : : : : :
934 GAATCTGTTTA.....ACTGTAAAGATCAGATGCGCAAGATGTGT 977

301 I.....SerProGluGlyIleGluPheValT 310
   : : : : : : : : : : : : : : : : : : : : : : : : : :
978 AGTGCTTAAGTAGAATCTACTTCTTAATCTATATATGTTGTGATTT 1027

310 hrProAlaIleGluLysIleAsnAlaLysGlyGluIleThrLeuAlaLys 326
   : : : : : : : : : : : : : : : : : : : : : : : : : :
1028 CAAGCGCGAAGTATCAGATCTGCTGCTAAAGTAAACGCTGAACCG 1077

327 GlyThrSerThrThrValLysAlaValTyLysLysAspGlyLysValVa 343
   : : : : : : : : : : : : : : : : : : : : : : : : : :
1078 GGAACAGCTGATGTACTGCAAAAGTTACATTACAGATGAGTGTGACT 1127

343 lAlaGluSer.....LysGluValLysValSerAlaGluG 355
   : : : : : : : : : : : : : : : : : : : : : : : : : :
1128 AACAAATACATTTAAAGTAGACAGTTACAGAACTGCTTCAAGTCAAA 1177

355 lYAlaAlaValAlaSerIleSerAsnTrpThrValAlaGluGln..... 369
   : : : : : : : : : : : : : : : : : : : : : : : : : :
1178 ATCAAGATTTACTTTAGTGAATCTTTTAATGCTTCAAGATACAA 1227

370 .....AsnLysAlaAspPheThrSerLysAspPheLysGlnAsn 383
   : : : : : : : : : : : : : : : : : : : : : : : : : :
1228 GTTGCATTTACAAAGCTGAGAAAGTAATCTCAATGTTGCTGGAAGAC 1277

383 nLys.....ValTyGlu.....GlyAspAsnAlaTyValG 394
   : : : : : : : : : : : : : : : : : : : : : : : : : :
1278 TAAACAGTGTCAATGTATGATACATAAAAGCGTGATCCTGAACATAAC 1327

394 lValAlaGluLeuLysAspGlnPheAsnAlaValThrThrGlyLysValGlu 410
   : : : : : : : : : : : : : : : : : : : : : : : : : :
1328 CTGTTGATTTCAAAAGT.....GCAACT 1350

411 TyGluSerLeuAsnThrGluValAlaValAlaLysAlaThrGlyLys 427
   : : : : : : : : : : : : : : : : : : : : : : : : : :
1351 GTACGTTCAATTAAATCCAAATTATTCACACAGCTGCTAATTAAGTAAGCA 1400

427 sValThrValLeuSer.....AlaGlyLysAlaProValLysV 440
   : : : : : : : : : : : : : : : : : : : : : : : : : :
1401 GCTCCTGTACAGCTAATGCTGCGCAATCTGGAAGACCTTCATTGAAG 1450

440 aThrValLysAspSerLysGlyLysAlaLeuValSerHisThrValGlu 456
   : : : : : : : : : : : : : : : : : : : : : : : : : :
1451 TAACATTTAAAGATTAAT.....ACAAAAAGAACATTTTACA 1485

457 lIleGluAlaPheAlaGlnLysAlaMetLysAspIleLysLeuGluLysTh 473
   : : : : : : : : : : : : : : : : : : : : : : : : : :
1486 GTTGATGTAAAAAAGACCTGTATTACAGATTAATAAGTAGAGCAAC 1535

473 rAsnValAlaLeuSer.....ThrLysAspV 482
   : : : : : : : : : : : : : : : : : : : : : : : : : :
1536 TTTCTTTAACTTTCGATGAAGCTGTGGCGGCGGGAAGTTGAAGAG 1585

482 aThrAspLeuLysValLysAlaProValLeuAspGlnTyGlyLysGlu 498
   : : : : : : : : : : : : : : : : : : : : : : : : : :
1586 TTAACCAAAAAACGATTAAGTAGTGAGTTCACCAATACGTAAGAA 1635

499 Phe.....ThrAlaPr 502
   : : : : : : : : : : : : : : : : : : : : : : : : : :
1636 ATTAATTTGCTACAAAAGTAAAGTTACTGTACAACTAATACAGAAAG 1685

502 ovalThrValLysValLeuAspLysAsp.....G 512
   : : : : : : : : : : : : : : : : : : : : : : : : : :
1686 ACTAGTTATTTAAATATGAATACGATTAATCAATTGACTTTGATACG 1735

512 lYLysGluLeuLysGluGln..... 518

```



```
|||||:|:|:|
1736 GCATFAGTGCACACCTGACCAATTTGTCGTCGACACAAAGACAAATT 1785
519 .....LysLeuGluAlaLysTyrValAsnArgGluLeuValLeuAs 532
1786 GTCATGCTGAAGTAGAGTAAATATTTCAAAAT ..... 1821
532 nAlaAlaGlyGlnGluAlaGlyAsnTyrThrValValLeuThrAlaLys 549
1822 ....GCTAGTACACACACCACTCAATCAATAAACATTTACTGTTATG 1867
549 ergLysGluLysGluAlaLysAlaThr...LeuAlaLeuGluLeuLysAla 564
1868 TAGTAAATGTAAGAGCTGACCTACACCAAGTAGATTTAGATTTGTACCA 1917
565 ProGlyAlaPheSerLysPheGluValArgGlyLeuAspThrGluLeuAs 581
1918 CCT.....TCTAAATATGATCTAAATGCTCCAAACACT ..... 1950
581 pLysTyrValThrGluGluAsnGlnLysAsnAlaMetThrValSerValL 598
1951 .GCTTCTACTGCAGATCTTGATTTATTAATTTGCAAGCTGTGAGATT 1999
598 euproValAspAlaAsnGlyLeuValLeuLysGlyAlaGluAlaGlu 614
2000 ACACACTGATTCAAATGCT .....AGACGTCAA 2028
615 LeuLysValThrThrThr .....AsnLysGluGlyL 625
2029 AAAAAAGTACTCCACACTGCACACTGCTAGCTACAAAAAATAAAAA 2078
625 sGluValAspAlaThrAspAlaGlnValThrValGlnAsnAsnSerValI 642
2079 AAAAGTTATGGAATGTTATTAACA .....TTCAAGGCAAGCAAGAT 2122
642 LeThrValGlyGlnGlyAlaLysAlaGlyGluThrTyrLysValThrVal 658
2123 TAAGCTATCACTTCTCTAGTACAGAAC ..... 2154
659 ValLeuAspGlyLysLeuLeuThrThrIserPheLysValValAspThr 675
2155 ...GTAGATGCA ..... 2163
675 rAlaProThrAlaLysGlyLeuAlaValGluPheThrSerThrSerLeuL 692
2164 .....ACAGCAGAGAAGATGACA .....A 2182
692 ySGluValAlaProAsnAlaAspLeuLysAlaAlaLeuLeuAsnIleLeu 708
2183 AACGTAAT...CCAGGAAATATATCAACTGCA ..... 2214
709 SerValAspGlyValProAlaThrThrAlaLysAlaThrAlaSer...As 724
2215 .....AGTGTACTGCGCAGTGCACACAGTACACACAGTCCGTGTAC 2255
724 nValGluPheValSerAlaAspThrAsnValValAlaGluAsnGlyThrV 741
2256 TGTAAAGCTTATTCAGTGAATGATTTAACATTTGAAGAATTAAATAT 2305
741 aLgAlaLysGlyAlaThrSerIleTyrValLysAsn ..... 753
2306 TCGGCTGTAATGACCTACACATTA...GTCAAAGATGACACATCAAC 2352
754 .....LeuThrValValLysAspLysGlyGlnGluLysValGluPheAs 768
2353 GAATTTATTTGCGTTCAAAACCGGCTAAAAATGATGATTTGTTATATA 2402
768 pLysAlaValAlaGlnValAlaValSerIleLysGluAla 780
2403 TAAACCGCTT .....GTAAACGTTAAAGATCA 2430
seq_name: gb_pat:AX000222
```

```
seq documentation block:
LOCUS AX000222 2763 bp DNA PAT 10-MAR-2000
DEFINITION Sequence 5 from Patent WO906567.
ACCESSION AX000222
VERSION AX000222.1 GI:7240686
KEYWORDS
SOURCE
ORGANISM Bacillus stearothermophilus.
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Geobacillus.
1 (bases 1 to 2763)
REFERENCE
AUTHORS Lubitz W. and Resch S.
TITLE SECRETION OF CARRIER-BONDED PROTEINS INTO THE PERIPLASMA AND THE
EXTRACELLULAR SPACE
JOURNAL Patent: WO 9006567-A 5 11-FEB-1999;
LUBITZ WERNER (AT); RESCH STEPHANIE (DE)
FEATURES
source location/Qualifiers
1..2763
/organism="Bacillus stearothermophilus"
/strain="pV72"
/db_xref="taxon:1422"
/clone="SBSB"
1..93
sig_peptide
1..2763
/note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAB7070.1"
/db_xref="GI:7240687"
/translation="MAYPKSYRKFPATATATAAMVAVPVVASAFDVAPOYKDA
IDFLVSTGATKGTERTKREYDEITRLDAVILARLKLIDVNAKDAGTDPKPKDA
YVNALEAGVLNGKAPKRGAYDPLTRVMAKITARVLIKADVKLPETDVNDWAP
VYKALKYEVGTGPTSPGAYQNTIRGDFGVYRAVNIADVPELVEYTAASITVK
VFNEDQIDAVDETFNFIIDNGLVTKATLSRDKSVYVNVNKEFTTRQETITATGINK
LKEGTPELTKGFVMSVDQAVTVALNNSLSLKYEESEGLKYKDDGKDVYGAVELTSS
NNIVAVSSGEVSVSAKYTAVPGRADYATVLPDGVLTNTEKVTVEVPVOYON
OGFTTLVDNLSNAPONTVAPNKAKEYTSMRAGERTYAMDTKNGDETPVPFKDATV
KSLNPLIATPAALNGSELVTANAGSGKASFETEYDNTNKRFTVDKPDVLODIK
DATSVKLSDEAVGGEGEVNOKTIKVASVDOYGEIKRGTGKVTVDKPDVLODIK
VNSDNTIDPDSGNSATDOFVVVATKDKIYNGKVEYVFNASDPTSTRTITVNVN
VKADATPVGDIAPSEIDVNAENFASADVDFINESVEIYTLDSNGRLKVPETA
TRLVGNTRYEVVNGVNLQFKNDELFTLSSSTVNVADVADGITRILPYKYNISASVP
ASRTVATSPVTVKYNLSDNDLPEELIRGVIDPTOLVXDNDINEPIAVSKAKNGCL
YKPLVTVKADASKVILPTGANNYGLHNDNTNGINPDEROAGLAKFFSVYHFDVPSL
ANYVTKSGTSVSSPSLSDAIDLTNSGDAVSFTLVKISLYVVGADKDNLLAAPVS
NVYVTK"
mat_peptide 94..2760
BASE COUNT 972 a 463 c 538 g 790 t
ORIGIN
alignment_scores:
Quality: 415.00 Length: 846
Ratio: 0.918 Gaps: 34
Percent Similarity: 53.428 Percent Identity: 23.995
alignment_block:
US-09-754-947-1 x AX000222 ..
Align seg 1/1 to: AX000222 from: 1 to: 2763
4 ThrPheProAspValProAlaAspHisTrpGlyIleAspSerIleAsnTyr 20
97 AGCTTCACAGAGATGTTGCCGCCCAATAT .....AAAGATGCGATCGATT 140
20 rLeuValGluLysGlyAlaValAlaValAsnAspLysGlyMetPheGluP 37
141 CTYAGTATCACTGCTGCAACAAAGGTAAACAGAAACAAATAATTCGGCG 190
37 roGlyLysGluLeuThrArgAlaGluAlaAlaThrMetMetAlaGlnIle 53
191 TTTACGATGAATCACTGCTAGATGCGCAGTATTCTTGCAAGAGTA 240
```



```

1915 CCT.....TCGAATGATGATGATGCTCCAAACACT..... 1947
581 pLysytrValThrGluGluAsnGlnLysAsnAlaMetHValSerValL 598
1948 .GCTTCACTGAGATGTTGATTATTATTAATTCGAAGGTGTGAGATT 1996
598 euproValAspAlaAsnGlyLeuValLeuLysGlyAlaGluAlaGlu 614
1997 ATACACCTGATTCATAGTGAACCGCTTAAAGGTTACCTCAGCTGCA 2046
615 LeuLysValThrThrAsnLysGluGlyLysGluValAspAlaThrAs 631
2047 ACTACACTGTGAGTACTAAT.....GATTATGTTGAAGTAAATGCG 2087
631 PALAGlnAlaThrValGlnAsnAsnSerValIleThrValGlyGlnGly 648
2088 GAATGTATTACATTCAGAGGGTAACGATGATTAACGCTA.....T 2128
648 lAlysAlaGlyGluThrThrLysValThrValAlaLeuAspLysLeu 664
2129 TAACCTCTTCACTAGTACAGTAAACGTTGATGTAACAGCTGATGCA 2172
665 lIleThrThrHis...SerPheLysValAlaAspThrAlaProThrAla 680
2173 ATTACAAACGATATTCAGTAAATATATCACTGCAAGTGTACCTGC 2222
680 sGlyLeuAlaValGluPheThrSerThrSerLeuLysGluValAlaPro 697
2223 CAGTCAACAGTACGACACACAGCTCCTGACTGTTAAGCTTAATTCAG 2272
697 snAlaAspLeuLysAlaAlaLeuLeuAsnIleLeuSerValAspLys 713
2273 ATATGATTTTAACTTGAAGATTAATATATTCGTTAATTCAC..... 2316
714 ProIleThrThrAlaLysAlaThrAlaSerAsnValGluPheValSer 730
2317 CCTACACATTAAGTCAAGATGACAGACATCAAC...GAATTATTCAG 2363
730 aAspThrAsnValAlaAlaGluAsnGlyThrValGlyAlaLysGly 747
2364 T.....TCAAAGCGCTAAAGATGATG 2386
747 hrSerIleThrValLysAsnLeuThrValValLysAsp 759
2387 GATATTGTATTAATACCGCTTGAAGCTTAAGAT 2424

seq_name: gb_ba:BSSBSBGEN
seq_documentation_block:
LOCUS      BSSBSBGEN      2916 bp      DNA      BCT      01-APR-1997
DEFINITION B.stearothermophilus sbsb gene.
ACCESSION  X98095
VERSION    X98095.1 GI:1418403
KEYWORDS   S-layer; sbsb gene.
SOURCE      Bacillus stearothermophilus.
ORGANISM   Bacillus stearothermophilus
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/staphylococcus group; Geobacillus.
            1 (bases 1 to 2916)
REFERENCE  Kuen,B., Koch,A., Asenbauer,F., Sara,M. and Lubitz,W.
            Molecular characterization of the Bacillus stearothermophilus PV72
            S-layer gene sbsb induced by oxidative stress
            J. Bacteriol. 179 (5), 1664-1670 (1997)
            97197531
JOURNAL    Koch,A.
            Direct Submission
            Submitted (30-APR-1996) A. Koch, Institute of Microbiology and
            Genetics, University of Vienna, Dr. Bohrergasse 9, A-1030 Vienna,
            AUSTRIA
FEATURES
            source          1..2916
                        Location/Qualifiers

```

```

/organism="Bacillus stearothermophilus"
/strain="PV72"
/sub_strain="p2"
/db_xref="taxon:1422"
/lab_host="Escherichia coli"
46..2808
/gene="sbsb"
46..2808
/gene="sbsb"
/codon_start=1
/transl_table=11
/protein_id="CAA66724.1"
/db_xref="GI:1418404"
/translation="MAYOPKSYRKFVATTAATAAMASAVAPVVSASFVDAPOYKDA
IDELVSTGATFKETKFGYDEITRFLDAALILARLKDIDVAKAGLEPDVADKDAK
YNAVALVEAVLNGKARKGACGADPLRIVEMAKILARYLKADVDKDDVADDAAP
YKALKIKREYTKGKPTPSFGAIONITRGEPAQGVYRAVNNAPPELVEYTAANSTYK
VFENTQIADVDFTNFAIDNGLVTKATLSRDKSVBVVNKPETRNQETTTATGINK
LKEPTAKELTGKFWMSVODAVTVALNNSSLKVEESGLYKDDGDKGVGAKVELTSS
NTNIVVSSGEVSVSAKYAVAPGTADYATKYLTPDGLVLTETFKVYTVLEVQYON
OGFTLVNDINSAPONTVAFNKAKEKVTSMFAGETKTYAMDTKNGDETPVDEKDATV
RSNPILITATAIINGSELLITANAGOSGKASFVTPKDNTRKPTTVKRDVADIDIKY
DATSVKLSDBANGGEGVEGYNOKRTIKVSAVDGYGKIKETKGTVTITRTEGLYIKN
VNSDNTIIDDSGNSATDQFVYVATKDKLVNGKVEYVEFNASDTPTSTKTLTVNVN
VKADATPVGLDIAVPEIDVNAENATSTADVDFTNESEYIYTLDSNGRLKRVPTA
TTLVAGNDVYEVNGVNLQFKNDELFTLSSSTVANDVADGITKRIYKIKASAVP
ASATVATSPVPTVYITLNSDNDLIFPEELIFGVITDPTOLVDEIDINEIFAYSKANDGLY
YKPELTVKVDASGKVIPTGANVYGLHNDATNGIMDEFQAGIAKFEFVHPVDLSL
ANVYKGTGSGTVSSPSLSDAIQLTNSGDVAFPLVTKISLYKADKDDNNLLAAPVS
NNTVTK"
46..138
/gene="sbsb"
139..2805
/gene="sbsb"
/evidence="experimental
/product="unnamed"
2847..2884
terminator
BASE COUNT 1019 a 485 c 572 g 840 t
ORIGIN

alignment_scores:
Quality: 415.00 Length: 846
Ratio: 0.918 Gaps: 34
Percent Similarity: 53.428 Percent Identity: 23.995

alignment_block:
US-09-754-947-1 x BSSBSBGEN ..
Align seg 1/1 to: BSSBSBGEN from: 1 to: 2916
4 ThrPheProAspValProAlaAspHisTrpGlyIleAspSerIleAsnTy 20
:::||||| ::::: ||:::|||||:::
142 ACCTTCACAGATGTTGCCCGCAATAT.....AAGATGCATGATT 185
20 lIeuValGluLysGlyAlaValLysGlyAsnAspLysGlyMetPheGluP 37
:::|||||::: ||:::|||||::: ::::: |||
186 CTACTATCAACACTGCTGCACAAAGTAAACAGAAACAAATTCGCGC 235
236 TTTACGATGAATCACTGCTAGATGGCGAGATTATCTTGCAAGACTA 285
37 roGlyLysGluLeuThrArgAlaGluAlaIleThrMetMetLacIle 53
|||||::: ||:::|||||::: ::::: |||
236 TTTACGATGAATCACTGCTAGATGGCGAGATTATCTTGCAAGACTA 285
54 lIeuAsnLeuProIleAspLysAspAlaLysProSerPheAlaAspSerG 70
|||||::: ||:::|||||::: ::::: |||
286 TTAACACTAGACGTTGACACCAAGACGAGCTTCACAGATGTGCC 385
70 nGlyIleThrPtyThrProPheIleAlaValGluLysAlaGlyValI 87
::: ::::: ||:::|||||:::
336 AAAAGAC...CGTGCAAATACGTCAACGCGCTTGAAGCTGCGTAT 382
87 lIeLysGlyThrGlyAsnGly...PheGluProAsnGlyLysIleAspArg 102

```

1220 ATCAAAGATTTACTTACTGATGATTAATCTTTCAATGCTCCACAGAATAACA 1265
370 AsnLysLlaaspherhserLysaspPheLysGlnasms 383
1270 GTTGCACTTTAACAAACCTGAGAAAGACTACTTCAATGTTGTCGAGAAAC 1315
383 nLys.....ValTyrGlu.....GlyaspasnaLaTyrValG 394
1320 TAAACAGTTGCAGTGTATGATGACTTAAACAGGTATCTCTGAACACTAAC 1365
394 InValGluLeuLysaspGlnPheasnalaValThrThrGlyLysValGlu 410
1370 CGGTGATTTCAAAGAT.....GCAACT 1392
411 TyrGlnSerLeuasnThrGluValAlaValaValaLysValaThrGly 427
1393 GTAGCTTCATTAATTCAAATATTTGCAACAGCTGCTATTAAATGATGAGA 1442
427 sValThrValLeuSer.....AlaGlyLysAlaProValLysV 440
1443 GCTCCTTGACACAGCTAATGCTGCCAATCTGAAAAGCTTATTGTAAG 1499
440 aThrValLysaspSerLysGlyLysAlaLeuValSerHisThrValGlu 456
1493 TTAACATTTTAAAGATAT.....ACAAAAGAACATTTACA 1522
457 IlegLualPheAlaGlnLysLametylsapIleLysleuGluLysTh 473
1528 GTTGAATGTCAAAAAGACCCTGATTAACAGATTTAAAGATGAAGAAC 1577
473 rAsnValaLeuSer.....ThrLysaspV 482
1578 TTCTGTAAACTTTCGATGAAGCTGTTGGCGGCGGAGTTGAAGAG 1622
482 aThrAspLeuLysValLysAlaProValLeuaspGlnTyrGlyLysGlu 498
1628 TTAACCAAAAAACGTTTAAAGTAAGTACGATGACCAATACGGTAAAGAA 1677
499 Phe.....Thalapr 502
1678 ATTAAATTTGGTACAAAAGTTAAAGTACTGTTACAACTAATACAGAGG 1727
502 ovalThrValLysValLeuAspLysasp.....G 512
1728 ACTACTTTATTAATAATGTAAATACGATATATACATTTACTTTGATGCG 1777
512 LysLysGluLeuLysGluGln..... 518
1778 GCAAATAGTCAAACTGACCAATTTGTTGTCGTTGCACAAAAGACAAATTT 1827
519LysLeuGlnAlaLysTyrValasnaTgGluLeuValLeuAs 532
1828 GTCAATGCTAAAGTGAAGGTTAAATATTTCAAAAT..... 1863
532 malaLaGlyGlnGlnLalagLysasTyrThrValValLeuThrAlaLys 549
1864 ...GCTAGTGACACACACACCAACTCTCAACTTAAACAATTTCTGTTAATG 1909
549 ergLysGlnLysGlnAlaLysAlaThr...LeuAlaLeuGluLeuLysAla 564
1910 TAGTGAATGTTAAAGCTACGCTACACAGATGATTAAGATATATCTGACA 1959
565 ProGlyAlaPheSerLysPheGluValaTargLysLeuaspThrGluLeuAs 581
1960 CCT.....TCGAAATTTGATGTGAATGCTCCAAACACT..... 1992
581 pLysTyrValThrGlnGlnGlnLysasnLysasnAlaMerThrValSerVal 598
1993 .GCTTCTACTGACAGATGTGTTATTTATTAATTTTGCAAAAGTGTGAGATTT 2041
598 eupProValaPalaAsnGlyLeuValLeuLysGlyAlaGlnAlaGlu 614
042 ATACACCTGATCTTAATGCTTAACGCTCTTAAAAAAGTTACTCCGAACGCA 2091

```
615 LeuLysValThrThrAsnLysGluGlyLysGluValAspAlaThrAs 631
      ::::: |||||
2092 ACTACACTTGTAGACTAAT.....GATATGTTGAAGTAATATGC 2132
      ::::: |||||
631 PAlaGlnValThrValGlnAsnSerValIleThrValGlyGlnGlyA 648
      ::::: |||||
2133 GAATGATTAATCAATTCAGGTAACGATGAATTAACGCTA.....T 2173
      ::::: |||||
648 IalysAlaGlyGluThrThrLysValThrValIleLeuAspGlyLysLeu 664
      ::::: |||||
2174 TAACTCTCTAGTACGTAACGTTGATGTAACGCTGATGCA..... 2217
      ::::: |||||
665 IleThrThrHis...SerPheLysValIleAspThrAlaProThrAlaLys 680
      ::::: |||||
2218 ATTACAAACGTAATTCAGTAATAATATATCAACTCTGCAAGCTGCTGCG 2267
      ::::: |||||
680 sGlyLeuAlaValGluPheThrSerThrSerLeuLysGluValAlaProA 697
      ::::: |||||
2268 CAGTGCACACAGTACGACACAGCTCTGTAAGCTTAATTCAGCTG 2317
      ::::: |||||
697 snAlaAspLeuLysAlaIleLeuLeuAsnIleLeuSerValAspGlyVal 713
      ::::: |||||
2318 ATATGATTTAATCAATTTGAAGATTAATATTCGCTAATTCAC..... 2361
      ::::: |||||
714 ProAlaThrThrAlaLysAlaThrAlaSerAsnValGluPheValSerAl 730
      ::::: |||||
2362 CCTACACAAATTAGTCAAAGATGACATCAAC...GAATTATTCAGCT 2408
      ::::: |||||
730 asPheThrAsnValIleAlaGluAsnGlyThrValGlyAlaLysGlyAlaT 747
      ::::: |||||
2409 T.....TCAAAGCGGCTAAATAAATGATG 2431
      ::::: |||||
747 hrSerIleThrValLysAsnLeuThrValIleLysAsp 759
      ::::: |||||
2432 GATATTTGTAATTAATTAACCGCTGTGTAACGTTAAAGAT 2469
      ::::: |||||
seq_name: gb_pat:AX033305
seq_documentation_block:
LOCUS AX033305 2766 bp DNA PAT 21-SEP-2000
DEFINITION Sequence 3 from Patent DE19903345.
ACCESSION AX033305
VERSION AX033305.1 GI:10280105
KEYWORDS
SOURCE
ORGANISM Bacillus stearothermophilus.
Bacillus stearothermophilus
Bacteria: Firmicutes: Bacillus/Clostridium group:
Bacillus/Staphylococcus group; Geobacillus.
1 (bases 1 to 2766)
REFERENCE
AUTHORS Lubitz,W.
JOURNAL Patent: DE 19903345-A 3 03-AUG-2000;
LUBITZ WERNER (AT)
FEATURES
source
1..2766
/organism="Bacillus stearothermophilus"
/db_xref="taxon:1422"
1..93
/feature="GI:10280106"
/feature="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC09983.1"
/db_xref="GI:10280106"
/feature="GI:10280106"
/feature="unnamed protein product"
sig_peptide
CDS
```

```
VDATSVKLSDEAVGGEVGVNOKTIRKVSANDVOYKIEFETKRAVTVTTNBEGLV
NNSDMTIDFSGNSATDQFVVVATKDKLVNGVEVKYKNNSDTPTSTKITVNVV
NKADATPVGLDIVAPSEIDVNAIPNTADVPFIMFESVEYITLDSNRLKVTPT
ATTLVGTNDEVNGNVLQFQKDELTLLTSSVAVDVADITRIPIVKYINSASV
PASATVATSPVTKLNSNDLTFEELIGVIDPQLVDEIDNEFIKVSAAKNDGY
LYNKPLVTVKDSAGKIIFPGANYVGINHATNTNITFDEQAGLAKKFSDFHFDVFS
LANVTVKSGTIVSSPSLSDAIOLTNSGAVSTLVTKSIYKAGADKDNLLAIVS
LVNVTYK"
mat_peptide 94..2763
BASP COUNT 973 a 464 c 537 g 792 t
ORIGIN
alignment_scores:
Quality: 413.50 Length: 846
Ratio: 0.913 Gaps: 35
Percent Similarity: 53.546 Percent Identity: 24.350
alignment_block:
US-09-754-947-1 x AX033305
Align seg 1/1 to: AX033305 from: 1 to: 2766
4 ThrPheProAspValProAlaAspHisTrpGlyIleAspSerIleAsnTy 20
::: ||||| ::::: |||||
97 AGCTTCACAGATGTTGCCCGCAATAT.....AAAGTGCAGTCAATTT 140
::: ||||| ::::: |||||
20 rLeuValGluLysGlyAlaValLysGlyAsnAspLysGlyMetPheGluP 37
||||| ::::: |||||
141 CTTACTATCAACTGCTGCACAAAGGTAAACAGACAAATTCGCGC 190
::: ||||| ::::: |||||
37 roGlyLysGluLeuThrArgAlaGluAlaIleThrMetAlaGlnIle 53
||||| ::::: |||||
191 TTACAGATGAATCACTCGCTAGATCGCGCATATCTTGCAAGAGTA 240
::: ||||| ::::: |||||
54 LeuAsnLeuProIleAspLysAspAlaLysProSerPheAlaAspSerGl 70
||||| ::::: |||||
241 TTAACATTAAGAGTGTGACACGCAAAAGACGACGCTTCACAGATGTC 290
::: ||||| ::::: |||||
70 nGlyIntTrpTyThrProPheIleAlaIleValGluLysAlaGlyValI 87
::: ::::: |||||
291 AAAGAC...CGTCAAAATACGTCAACGCCCTGTGTAACGTGCGAT 337
::: ||||| ::::: |||||
87 IeLysGlyThrGlyAsnGly...PheGluProAsnGlyLysIleAspArg 102
::: ||||| ::::: |||||
338 TAAACGTAAAGACACCTGCAAAATTTGGTCATACGACCATTAACTCGC 387
::: ||||| ::::: |||||
103 ValSerMetAlaSerLeuLeuValGluAlaIleThrLysLeuAsp..... 116
388 GTTGAATGCAAAATATCATCGCAACCGTTACAAATTAAGCTGACGA 437
::: ||||| ::::: |||||
117 .....ThrLysValAsnGlyThrProAlaThrLysPheL 128
||| ||||| ||| |||
438 TGTAAACTTCCTCATTCACCTGATGATTAACGATACAGGACCATACGTA 487
::: ||||| ::::: |||||
128 ysAspLeuGluThrLeuAsnTrpGlyLysGluLysAlaAsnIleLeuVal 144
||| ||| ::::: |||||
488 AAGCGCTT.....TATAATACGAAGTAAACCAAAAGTTAAAA 525
::: ||||| ::::: |||||
145 GluLeuGlyIleSerValGlyThrGlyAspGlnTrpGluProLysLysTh 161
::: ||||| |||
526 CACCAACACAGCTTCGTCATAC.....AAAAA 554
::: ||||| ::::: |||||
161 rValThrLysAlaGluAlaIleGlnPheIleAlaLysThrAspLysGluP 178
::: ||||| ::::: |||||
555 CATCACTTCCTGCTGACTTGGCGCAATTTGTAATACAGCGGTGAATAT 603
::: ||||| ::::: |||||
178 heGlyThrGluAlaIleLysValGluSerAlaLysAlaValThrThrGln 194
::: ::::: |||||
604 ....AATGACAGTGCAGAAATAGTGAAGTACGCGGTTAATTCGACT 648
::: ||||| ::::: |||||
195 LysValGluValLysPheSerLysAlaValGluLysLeuThrLysGluAs 211
```

```
649 ACAGTGAAGTAACTCAATACGCAAAATT.....GCTGA 683
211 pIleuysValThrAsnLysAlaAsnAsnAspLysValLeuValLysGluV 228
684 TGTGATTTACACAAATTTTGTCTATGATACGTTAACTGTTACTAAAG 733
228 aLThrLeuSerGluAspLysArgSerAlaThrValGluLeuTyrSerAsn 244
734 CAACCTTTCTCGTGAATAAAATCCGTAGAGTGTGTTAAATAAACC 783
245 LeuAlaLeuLysGluThrTyrThrValAspValAsnLysValGly.... 259
784 TTTACTCGTAATCAGGAATATACAAATTACAGCAGCGCATTAATAATT 833
260 LysThrGluValAla.....ValGlySerLeu. 268
834 AAAGCGGACCGCTAAGCAATTAAGTAACTGTTGTTGCTGCTGC 883
269 ..GluAlaLysThrIleGluMetAlaAspGluThrValAlaAspGlu 284
884 AAGATCCGCTAACTGTCACATAAATAGTTCGCTAAAGTTGGAGAG 933
285 ProThrAlaLeuGluPheThrValLysAspGluAsnGlyThrGluValVa 301
934 GAATCTGGTTTA.....ACTGTAAGAATCAGAGTGCAGAAAGATCTGT 977
301 L.....SerProGluGlyIleGluPheValT 310
978 AGGTGCTAAAGTAGAAGTAACTTCTTCTAATACTAATATATTGTTAGTT 1027
310 hProAlaAlaGluLysIleAsnAlaLysGlyGluIleThrLeuAlaLys 326
1028 CAAGTGGCCAGACTATCAGTATCTGCTGAAGTTACAGCTGTAAACCG 1077
327 GlyThrSerThrThrValLysAlaValTyrLysLysAspGlyLysValVa 343
1078 GGAACACGCTGATGTTACTGCAAAAGTTACATTACAGATGGTGTGACT 1127
343 lAlaGluSer.....LysGluValLysValSerAlaGluG 355
1128 AACAAATACATTTAAAGTCAGACAGTACAGAACTGCCTGCAGAAACAA 1177
355 lAlaAlaValAlaSerIleSerAsnTyrPheValAlaGluGln..... 369
1178 ATCAAGGATTTACTTAACTTGAATCTTCTAATCTCCACAGAAATACA 1227
370 .....AsnLysAlaAspPheThrSerLysAspPheLysGluAsnAs 383
1228 GTTGCATTTAACAAAGCTGAGAAAGTAACTTCAATGTTTGTGAGAAAC 1277
383 nLys.....ValTyrGlu.....GlyAspAsnAlaTyrValG 394
1278 TAAACACATGTCATGTATGATACATAAAACGGTGATCCTGAACATAAC 1327
394 lValAlaGluLeuLysAspGlnPheAsnAlaValThrThrGlyLysValGlu 410
1328 CTGTTGATTTCAAAAGT.....GCAACT 1350
411 TyrGluSerLeuAsnThrGluValAlaValAlaAspLysAlaThrGlyL 427
1351 GTACGTTATTAAATCAATTAATTCACACAGCTGCTAATTAATGTAAGGA 1400
427 sValThrValLeuSer.....AlaGlyLysAlaProValLysV 440
1401 GCTCCTTCTCACAGCTAATGCTGCCAATCTGGAAGAGCTTCATTGGAAG 1450
440 aLThrValLysAspSerLysGlyLysAlaLeuValSerHisThrValGlu 456
1451 TAACATTTAAAGATTAAT.....ACAAAGAAACATTTTACA 1485
457 lIleGluAlaPheAlaGlnLysAlaMetLysAspIleLysLeuGluLysTh 473
.....
```

```
1486 GTTGATGTGAAAAAGACCCTGTATTACAAAGATATTAAAGTAGATCAAC 1535
473 rAsnValAlaLeuSer.....ThrLysAspV 482
1536 TTCTGTTAACTTTCCGATGAAGCTGTGGCGGGGGAAGTTGAAGAG 1585
482 aLThrAspLeuLysValLysAlaProValLeuAspGlnTyrGlyLysGlu 498
1586 TTAACCAAAAAAGATTAAGTAAGTGCAGTTGACCAATACGTAAGA 1635
499 phe.....ThrAlaPr 502
1636 ATTAATTTGTGACAAAGTAAGTAACTTACTTACAACTAATATACAGAG 1685
502 oValThrValLysValLeuAspLysAsp.....G 512
1686 ACTAGTATTAAATAATGTAATAGCATTAATCAATTGACTTTGATAGCG 1735
512 lLysGluLeuLysGluGln..... 518
1736 GCAATAGTCAACAGTACCAATTTGTTGCTTGCAACAAAGACAAAAATT 1785
519 .....LysLeuGluAlaLysTyrValAsnArgGluLeuValLeuAs 532
1786 GTCAATGTAAAGTAGAAGTAAATATTTCAAAAAT..... 1821
532 nAlaAlaGlyGlnGluAlaGlyAsnTyrThrValValLeuThrAlaLys 549
1822 ...GCTAGTGCACACACCAACTTCACTAAACAAATTAAGTTAAG 1867
549 eGlyGluLysGluAlaLysAlaThr...LeuAlaLeuGluLeuLysAla 564
1868 TAGTAATGTAAAGCTGACCTGACACAGTAGATTAATGATTGTAGACA 1917
565 ProGlyAlaPheSerLysPheGluValArgLysLeuAspThrGluLeuAs 581
1918 CCT.....TCTGAATTTGATGTAATGCTCCAAACACT..... 1950
581 pLysTyrValThrGluGluAsnGlnLysAsnAlaMetThrValSerVal 598
1951 .GCTTCTACTAGACAGATGTTGATTATTAATTTCCAAAGTGTGAGATT 1999
598 euProValAspAlaAsnGlyLeuValLeuLysGlyAlaGluAlaGlu 614
2000 ATACACATGATTTAATGATTAACCGCTTAAAAAAGTTACTCCACTGCA 2049
615 LeuLysValThrThrThrAsnLysGluLysGluValAspAlaThrAs 631
2050 ACTACACTGTAGTACTAAT.....GATTATGTTGAAGTTAAATGG 2090
631 pAlaGlnValThrValGlnAsnAsnSerValIleThrValGlyGlnGlyA 648
2091 GAATGATATTACAAATTCAGGGTAACAGATGAATTAACGGCTA.....T 2131
648 lLysAlaGlyGluThrTyrLysValThrValValLeuAspGlyLysLeu 664
2132 TAACCTCTTCACTAGTACAGTAAACGTTGATGTACAGCTGATGAGA.... 2175
665 lIleThrThrHis...SerPheLysValAlaAspThrAlaProThrAlaL 680
2176 ATTACAAAACGATTTCCAGTAATAATATATCACTGTGCAAGTGAACGTC 2225
680 sGlyLeuAlaValGluPheThrSerThrSerLeuLysGluValAlaProA 697
2226 CAGTGCACACATGACACACAGTCCGTTACGTTAAGCTTAATTAATG 2275
697 sAlaAspLeuLysAlaAlaLeuLeuAsnIleLeuSerValAspGlyVal 713
2276 ATAATGATTAAACATTTGAAGAAATTAATATTCGGTAAATGAC..... 2319
714 ProAlaThrThrAlaLysAlaThrAlaSerAsnValGluPheValSerAl 730
2320 CCTACACAAATTAATGCAAGATGAAGACATCAAC...GAATTTATTTGCACT 2366
```

```
730 aaspthrsnvalvalaaglaanglythrvalglalalysgllyalat 747
      :          :          :          :          :
2367 T.....TCAAAAGCGGCTAAAGATGATG 2389

747 hserllevallysasnleuthrvalvallysasp 759
      :          :          :          :          :
2390 GATATTGTATATAAACCGCTTAAAGAT 2427

seq_name: gb_ba:BACSLPZ

seq_documentation_block: 4251 bp DNA BCT 03-OCT-1994
LOCUS BACSLPZ
DEFINITION Bacillus sphaericus 2362 (clone pgC4.2) surface-layer protein gene,
complete cds.
ACCESSION M28361
VERSION M28361.1 GI:341911
KEYWORDS surface layer protein.
SOURCE Bacillus sphaericus (strain 2362) DNA.
ORGANISM Bacillus sphaericus
          Bacillus/staphylococcus group; Bacillus.
          1 (bases 1 to 4251)
REFERENCE Bowditch, R.D., Baumann, P. and Yousten, A.A.
TITLE Cloning and sequencing of the gene encoding a 125-kilodalton
surface-layer protein from Bacillus sphaericus 2362 and of a
related cryptic gene
JOURNAL J. Bacteriol. 171 (8), 4178-4188 (1989)
MEDLINE 89327128
FEATURES
    source
        1..4251
            /organism="Bacillus sphaericus"
            /strain="2362"
            /db_xref="taxon:1421"
        60..64
            /-35_signal
        84..89
            /-10_signal
        164..168
            RBS
        178..3708
            CDS
            /codon_start=1
            /transl_table=11
            /product="surface layer protein"
            /protein_id="AA50256.1"
            /db_xref="GI:556012"
            /translation="MAKONKGRKFFAASATAIIVPASAOLINDENKISGVAK
            EAYOSLIDAGVIOGDANGFNPLKTIISRAEATITFTALIEAGDVFNDVKADAMY
            YDAIAATVENGIEFGVATGEPAFKOLTRSEAKIIVDAELEGDLSERADASTVK
            PMAKSYLEIVANGVITGSEANGKTNLNPAPITROPFAVFSRTINVAATPVDKI
            EVDAKLTLEISDGTKEVYLEKALPEPKNETEYTRIKOYEAKTYIVVTTATK
            SVSATNKEVYVEPFGTVDEKETAEDANVIALKSKTKSVSLADNKATVITLTKLN
            NNRADALISINVAADKEIIVKNVEFAVDNKIPEVEIVSLGTRAKAVLSEVENL
            SSTNFTLDGRKAVGNNVVGAGNKTVILTPYSSALSGDKLTVSGAKDFAGVSLNS
            THEEKVVEDEAPTEATATLETVTLFSESDIMDTPVSKSNVWKGGDSKESESE
            RIDNKKKPEVGESEKTLPGKVDVYVEDIKDYSDNKLADTKVITPEIDOTPEVR
            KVALDEKTIKVEKSTVDDESAIKTGNVYKDDKDVYSDKTVVSKSKSYITDL
            YSKVSGEENTITTKNVADATRLNNTMLDYGKFTRSKEGDEHVNADKAKVVL
            KFDKMDAASLADSNLVKINDPLQSLSEDAVLVSNDATVVTITFAETIKAKVVL
            PASGKAISGSKVNVNLOWGVATGSGNVHKKNGSEENKITLSTSTPLKALIDKD
            YDAKYTAELVDKRTKVKFSTVINSAAANFTSEHKSIDSIOVNGSTVYKFEDEIN
            TMSADKLKVLKSLVDIAGNESNTNPDIKAGINDLSVAIPVAVGEVPYDKETIE
            TFSNLTDSVSGEVLSPDPTVRYSDNKLAIKRYSAIANNNNVVITLSDNREVAA
            YKTRAKNAKLITDNDGCKNAIDFTTKTKATKVASSTLSLDAKTMNNETTRAKA
            KATGTEGTAATNOIVGSKDALQVAIDVAELVKNPAAATLOOLDAKTDLTAITAYNA
            AKVEDISLLVAPDLVLTGTDNGTITGFAVGTGTLKVTSDSAANVEVTDPTGLAVTA
            KAKGEANILVOLTGKDKVIRKGVKVTYVSE"
    terminator
        3764..3806
            /note="putative"
BASE COUNT 1552 a 681 c 781 g 1237 t
ORIGIN
```

```
alignment_scores:
    Quality: 402.50
    Ratio: 0.833
    Length: 916
    Gaps: 36
```

```
Percent Similarity: 52.729 Percent Identity: 23.144
alignment_block:
US-09-754-947-1 x BACSLPZ ..
Align seg 1/1 to: BACSLPZ from: 1 to: 4251

16 AspSerlleasntryleuvalglulysgllyalalalysgllyasnasp 32
   :          :          :          :          :
310 GAAGCAGTTCAATCTTAGTACCGCTGGTAAATCCAAAGGAGGCTAA 359

32 sgllymetphegiuproglylysgluleuthrarglaaglaalalatr 49
   :          :          :          :          :
360 CGGCAACTTCAACCCCTAAAGCTATCTCAGTCCGGAGAGTGTCTACA 409

49 etmetalaglinleuasnleuproileasplyaspalalysproser 65
   :          :          :          :          :
410 TCTTCACATACCGCTCTAGAAATTAAGAGCAAGAGTGATGTA..... 453

66 phealaaspserglnglyln...trptrythrprophelialealalava 81
   ||  ||  :          :          :          :
454 TTCAAAGACGTTAAAGCGATGCTGTGATCTAGATGCTATCGCAGCAAC 503

81 lglulysalaglyvaliilelysgly...thrglysnnglyphegiupro 97
   :          :          :          :          :
504 TGTAGAAAACGGAAATTTTGAAGGTGTAAAGTCTACTGATTCGACACCA 553

97 snglylysleasparqvalsermetalaserleuvalglulalaty 113
   ||  :          :          :          :
554 ACAACAATTAATCTGCTCTGAGAGCTGCTAAATTTAGTATGATGCTTTC 603

114 lysleuaspthrlysalalasnnglythrproalathrlysphe 130
   :          :          :          :          :
604 GAATTAGAGCGTGAAGCGCAT.....CTAAGCAATTCGCTGACGC 644

130 ucluthrleuasn...trpglylysglulysalasnileuvalglul 146
   :          :          :          :          :
645 TTCTACTGTTAAACCATGGGCTTAATCTTACCTAGAAATCGCAGTTGCAA 694

146 euglylleservalglythrly..... 153
   ||  :          :          :          :
695 ACGCGCTTATCAAGGTCTGTGAAGCAATGTAAACCAACTTAACCCA 744

153 ..... 153

745 AATGCTCAATTACTCGCAAGACTTCGAGTGTATTCACGATCAT 794

154 ..... 157

795 TGAAGACGTAGATGCTACTCCAAAGTTGACAAATCGAAGTAGTTGACG 844

158 ..... 163

845 CTAAACCTTAAACGTTACTTATCTGACGCTACGTAAGAACTGTACT 894

164 lysalaglaalaglinphelealalalysrthrasplysglnphegly.. 179
   :          :          :          :          :
895 TAGAAGAAAGCTTAGAGCCCTAACAAGAAACAGAACTTACTTCAAAAT 944

180 ..... 183

945 TAAGCATGTTGAATACAAAGCTAAAGTTACTTATGTTGTTGACGTA 994

183 lalysvalgluaseralalalalalalthrthrglulysvalglulalys 199
   :          :          :          :          :
995 CTCACATTAATCTGATCTGACACATACCTTAAGAACTAGATGTA 1044

200 pneserlysalaalaglulysleuthrlysglulaspilalysvalthr 216
   ||  :          :          :          :
1045 TTCAGACGTACTGTTGATTAAGAAACAGCTGAAAGAT.....GCAAGCTAA 1088

216 nlysalasnasasplysalvalleuvallysglulalthrleuserglua 233
   |  ||  :          :          :          :

```

1089 CTACGCTTAAATCAGGTAACAATTAAATGTAATCTTACGCTG 1138
 233 sPLysaIrgSerAlaThrValGluLeuTySeraIleuAlaIalysGln 249
 ||| : : : : : ||| : : : : :
 1139 ATATATAACAGCTACTGTTACACTTAACGATTAACCTTAACACACAA 1188
 250 ThTThrValAlaSpValAsn.....LysValGlyLysThrGluValAl 264
 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 1189 GCGATGCTATTAGCATTTCTAATGTAAGCGCGCATTAAGAAATCAA 1238
 264 aValGlySerLeu..... 268
 ||| : : : : :
 1239 TGTGAAAAATGTGAATTACACAGCTAGATTAATAATCCGAAGTTA 1288
 269 ..GluAlaLysThrIleGluMetAlaAspGlnThrValAlaIalysGlu 284
 ||| : : : : : ||| : : : : :
 1289 CTGAAGTTAACTTATAGGAACAAAGCTGTAAAGTTACATTATCTGAA 1338
 285 Pro.....ThrAlaLeuGlnPheThrValLlysAspGlu..... 295
 ||| : : : : : ||| : : : : :
 1339 CCGTTGACAACTTAAGTCTTACAAACTTCACCTTAGACGGAAGCGTA 1388
 296AsnGlyThrGluValValSerP 303
 ||| ||| : : : : :
 1389 CTTGGGAATGTTGTTATGCGTCTGCTAATAAACAGTTATCTTAAC 1438
 303 roGluGlyIleGluPheValThrProAlaIalGlyLysIleAsnAlaLys 319
 || : : : : : ||| : : : : :
 1439 CTTACAGCTCTTCTCGCACTTCTGTCGGATCATTAAGCTTACTGTTCT 1488
 320 GlyGluIleThrLeuAlaLysGlyThrSerThrThrValLysAlaVal 336
 ||| : : : : : ||| : : : : :
 1489 GGTGCAAAAGACTTCGCTGTTTGTTCATTAATTAATCAACGCAAGT 1538
 336 rLysLysAspGlyLysValValAlaGluSerLysGluValLysValSer 353
 ||| : : : : : ||| : : : : :
 1539 TAA.....GTTCTGACATTAAGAAAGCAACCAACACTAA 1573
 353 IacGluGlyAlaValAlaSerIleSerAsnTrpThrValAlaGluGln 369
 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 1574 CTGAAGCTACTGCACCTTGAAACTGTATACATTACATTCACAGAAAT 1623
 370 AsnLysAlaAsp.....PheThrSerLysAs 378
 ||| : : : : :
 1624 ATTGACATGATCTCTAAAGCTTAAACGTTTATTGGAATCTGCA 1673
 378 PheLysGlnAsnAsnLysValTyrcL.....GlyAspAsnAlaTyrcV 393
 ||| : : : : : ||| : : : : :
 1674 TTCTAAGAAAGACATCTGAATTCGAGCGTATGCTGATTAATATACA 1723
 393 aGlnValGluLeuLysAspGlnPheAsnAlaValThrThrGlyLysVal 409
 ||| : : : : : ||| : : : : :
 1724 AATTCGTA...TTCAAAAGCTCTGAAAAACCTTCCAACTGGAAGTA 1770
 410 GluTyrc.....GluSerLeuAsnThrGluValAlaVala 420
 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 1771 GATGCTACGTAAGACATTAAAGTTACTCTGATACAAATGCTAA 1820
 420 lValAspLysAlaThrGlyLysValThrValLeuSerAlaGlyLysAlaP 437
 ||| : : : : : ||| : : : : :
 1821 ACATCAAAAGTTACTGTAACCTCGAATTCGACCAAACTCGTCCAGAG 1870
 437 roValLysValThrValLysAspSerLysGly...LysAlaLeuValSer 452
 ||| : : : : : ||| : : : : :
 1871 TAAAGAAATACTGCTCTTGTGATGAATAAATCATATCATCTCT 1920
 453 HlStrValGluIleGluAlaPheAlaGlnLysAlaMetLysAspIle 469
 ||| : : : : : ||| : : : : :
 1921 AAACTGTTGAGAGAGCTCTATAAAACAGCAACTACACAGTTAA 1970
 469 sLeuGluLysThrAsnValAlaLeuSerThrLysAspValThrAspLeu 486
 ||| : : : : : ||| : : : : :
 1971 AGACAAAGCAGCAAGTAGTTCTGTTGATTAAGTAACCTGTTGATCAA 2020

486 ys...ValLysAlaProValLeuAspGlnTyrcLysLysGluPheThrAla 501
 ||| : : : : : ||| : : : : :
 2021 AAGATCTTAATCTGTAATATTATGACCTATACCAAAAGTAAGTGTGGT 2070
 502 ProValThrValLysValLeuAsp...LysAspGlyLysGluLeuLysG 517
 ||| : : : : : ||| : : : : :
 2071 GAAAAATCAATTCAAATTAAGATGTTAAAGATGCAACAAACCTTAACA 2120
 517 uGlnLysLeuGluAlaLysTyrcValAsnTrpGluLeuValLeuAsnAla 534
 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 2121 TACATGCTAGAT.....TACACTGTAA.....TTTACAGAT 2155
 534 IacGlyGlnAlaGlyAsnTyrcThrValValLeuThrAlaLysSerGly 550
 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 2156 CAGATTAAGAGAGTCCAGATTAATGACAGATGATCAATGCTATGCA... 2202
 551 GluLysGluAlaLysAlaThrLeuAlaLeuGluLeuLysAlaProGlyAl 567
 ||| : : : : : ||| : : : : :
 2203 ...AAAGCTAAAAAGTTGTTTAAATTCGACAAAAAATGATGATCAGC 2249
 567 a.....PheSerLysPheGluValArgGlyLeuAspThr.... 578
 ||| : : : : : ||| : : : : :
 2250 ATCTTAGCTGACTATTCAACTACTAGTAAATCAATGATACTTTAC 2299
 579 ..GluLeuAspLysTyrcValThrGluGluAsnGlnLysAsnAlaMetThr 594
 ||| : : : : : ||| : : : : :
 2300 AAACATTAACAGAAAGATGTTGCTACACTTCAAGTTCAATGATGCTACA 2349
 595 ValSerValLeuPro.....ValAs 601
 ||| : : : : : ||| : : : : :
 2350 GTAGTACATTAATCTTTGACAGACAAATTAAGTGACGATGTTTAA 2399
 601 PalaAsnGlyLeuValLeuLysGlyAla.....GluAlaAlaGlu 615
 ||| : : : : : ||| : : : : :
 2400 TGCTTCTGTTAAAGCAATTCGATCTGTAAGTAAGTAAGTAAGAT 2449
 615 eulysVal.....ThrThrAsnLysGluGlyLysGluVal 627
 ||| : : : : : ||| : : : : :
 2450 TACAGTTATGGAGTAAGAAAGATCTTCTGTAACTGCAAGAAATTT 2499
 628 AspAlaThrAspAlaGlnValThrValGlnAsnAsnSerValIleThrVal 644
 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 2500 AATGTTCAGAAATTAATTAATCTTATCATCAACATCA.....ACGCC 2543
 644 lGlyGlnGlyAlaLysAlaGlyGluThrTyrc...LysValThrVal. 658
 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 2544 ATTAAGCTTCTGTAATAATGACAAAGATTAATGCAAAATACACAGCAG 2593
 659 ..ValLeuAspGlyLysLeuIleThrThrHisSerPheLysValAlaAsp 674
 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 2594 AATTAGTTATGAAAAACTGTAAAGTTAATTCCTACACTTATTAAC 2643
 675 ThrAlaProThrAlaLysGlyLeuAlaValGluPheThrSerThrSer 691
 ||| : : : : : ||| : : : : :
 2644 TCAGCA.....CTGCAAAATGCATTTACTTCTGAAATCA 2678
 691 uLys..... 692
 ||| : : : : :
 2679 TAAATTAATGATCTTCAAGTGAATGATCTCAACTGTATACAGTTAAT 2728
 693GluValAlaProAsnAla.....AspLeuLysAlaAla 703
 ||| : : : : : ||| : : : : :
 2729 TCAAGATGAAATTAACAAACGCTTCTGATTTGATTAAGTAAT 2778
 704 LeuLeuAsnIleLeuSerValAspGlyAlaProAlaThr.....ThrAl 718
 ||| : : : : : ||| : : : : :
 2779 TTATCTAATTAATGATTTGACAGAAACGAATACAAATTAATATCC 2828
 718 aLysAlaThrAlaSerAsnValGluPheValSerAlaAspThrAsnVal 735
 ||| : : : : : ||| : : : : :
 2829 AATGCCAATTAAGCTGATTAATTAATTTAGATTTCTGTCACAGTTG 2878

[illegible][illegible]


```

697  snnlaaspleuysalalaleuleu.....Asnleleu 708
      ||| : : : : : ||| : : : : :
3438  AATTTTCATTTAAACGATTCATTTTGGTAAAGCTGATGTCGACCACTTA 3487
      ||| : : : : : ||| : : : : :
709  SerValAspGlyValAlaProLathrThrAlaLysAlaThrAlaSerAsnVa 725
      ||| ||| : : : : : ||| ||| |||
3488  GTTAAAGACGAGTACGATTTTTCGGTGTCCCTTAAACGTGTCCAAACGG 3537
      ||| ||| : : : : : ||| ||| |||
725  IGIupheValSerAlaAspThrAsnValValAlaGluAsnGlyThrValG 742
      : : : : : : : : : : : : : : :
3538  TGGTTTACTTA...TACACAAACCCACTGATTCATCTATTACTGTGATCGT 3584
      : : : : : : : : : : : : : : :
742  IyAlaLysGlyAlaThrSerIleTyI 751
      ||| ||| ||| : : : : : ||| |||
3585  GTTAAAGCATTTGCAACTGTTGTTTAAACGTA 3613

```


OM of: US-09-754-947-1 to: EST: * out_format : pfs
Date: Mar 23, 2002 8:28 AM

About: Results were produced by the Gencore software, version 4.5.
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=fname.p2n.model -DEV=xlp
-O/cg22_1/USPTO.spool/US09754947/runal_22032002_125444_25431/app_query.fasta_1.854
-DB=EST -QMT=fastap -SUFFIX=est -GAPOP=12.000 -GAPEXT=4.000
-MIMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCLIN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USER=us09754947_rcgn1_1_4528 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPRY -WAIT -THREADS=1

Search information block:

Query: US-09-754-947-1

Query length: 785

Database: EST:*

Database sequences: 11351937

Database length: 1077921985

Search time (sec): 1536.890000

Score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_hlc:AK014540	+	172.00	267.76	9.5e-06	2880	AK014540 Mus musculus 0 day ne
gb_gss:A2538390	+	149.00	239.70	0.0003	886	A2538390 ENTREP05TR Entameoba h
gb_gss:BH135264	-	141.00	226.03	0.0020	899	BH135264 ENTREP45TR Entameoba h
gb_hlc:BC003268	+	141.00	214.68	0.0086	3084	BC003268 Mus musculus, Similar
gb_gss:CN056M47	+	140.50	223.83	0.0026	1042	AL405221 T7 end of clone A20A
gb_gss:BH135281	+	138.00	220.81	0.0039	913	BH135281 ENT0A30TR Entameoba h
gb_hlc:AK004594	+	138.00	210.19	0.0152	2893	AK004594 Mus musculus adult ma
gb_hlc:BC008524	-	137.00	204.68	0.0309	4379	BC008524 Mus musculus, Similar
gb_gss:A2538324	-	135.50	217.04	0.0063	869	A2538324 ENTREP05TR Entameoba h
gb_hlc:AK019969	+	135.00	206.79	0.0236	2413	AK019969 Mus musculus 8 days c
gb_hlc:AK019450	+	134.50	205.14	0.0291	2633	AK019450 Mus musculus 14, 17 d
gb_hlc:AK016539	+	133.00	202.78	0.0394	2382	AK016539 Mus musculus adult ma
gb_gss:A2674287	-	131.50	210.19	0.0152	877	A2674287 ENTREP33TR Entameoba h
gb_gss:CN0570GA	+	131.00	209.12	0.0166	943	AL423584 T3 end of clone AZ0A
gb_hlc:BC008697	+	131.00	201.49	0.0465	2056	BC008697 Homo sapiens, clone I
gb_hlc:AK004893	+	129.50	195.53	0.0998	2983	AK004893 Mus musculus adult ma
gb_hlc:AK004966	+	129.00	195.06	0.1061	2864	AK004966 Mus musculus adult ma
gb_gss:CN0709U	-	128.50	204.72	0.0307	915	AL423352 T3 end of clone AZ0A
gb_hlc:AK013048	+	128.50	197.82	0.0744	1936	AK013048 Mus musculus 10, 11 c
gb_hlc:BC011416	+	128.50	196.53	0.0878	2226	BC011416 Homo sapiens, Similar
gb_hlc:AK004862	+	127.50	192.11	0.1547	2993	AK004862 Mus musculus adult ma
gb_hlc:AK016519	+	127.00	188.19	0.2558	4581	AK016519 Mus musculus adult ma
gb_hlc:BC006984	+	127.00	197.05	0.0832	1598	BC006984 Homo sapiens, Similar
gb_hlc:AK017839	+	127.00	196.50	0.0861	1695	AK017839 Mus musculus 8 days g
gb_gss:CN056Z0T	-	125.50	199.17	0.0626	963	AL422667 T3 end of clone AZ0A
gb_hlc:AK014398	+	125.00	199.13	0.1995	2345	AK014398 Mus musculus adult ma
gb_hlc:BC011226	-	124.50	198.40	0.2492	2582	BC011226 Mus musculus, Similar
gb_gss:CN0600DE	-	122.50	194.04	0.1208	969	AL410520 T3 end of clone AZ0A
gb_gss:CN0570D0	+	121.50	191.06	0.1770	1114	AL409718 T3 end of clone XAVO
gb_hlc:AK015000	+	118.00	186.93	0.3006	1806	AK015000 Mus musculus adult ma
gb_hlc:BC011801	+	118.00	179.25	0.8057	2113	BC011801 Homo sapiens, clone 1
gb_hlc:BC003664	+	118.00	177.88	0.9601	2451	BC003664 Homo sapiens, Similar
gb_est1:AA264459	+	117.50	188.87	0.2345	678	AA264459 LD08971.Sprline LD Dros
gb_hlc:BC003279	+	117.50	178.55	0.8805	2078	BC003279 Mus musculus, Similar
gb_est2:BG638449	-	117.00	186.53	0.3164	797	BG638449 LD25659.Sprline LD Dros
gb_hlc:AK004574	+	117.00	174.42	1.50	2971	AK004574 Mus musculus adult ma

gb_gss:A2530606 + 116.50 185.18 0.3763 882 I A2530606 ENT0A11TR Entameoba
gb_gss:A2691742 - 116.50 184.49 0.4114 908 I A2691742 ENT0A66TR Entameoba
gb_gss:A2683303 - 116.00 183.59 0.4615 913 I A2683303 ENT0A21TR Entameoba
gb_hlc:AK011815 + 116.00 178.53 0.8828 1581 I AK011815 Mus musculus 10 da

seq_name: gb_hlc:AK014540

seq_documentation_block:

LOCUS AK014540 2880 bp mRNA HTG 05-JUL-2001
DEFINITION Mus musculus 0 day neonate skin cDNA, RIKEN full-length enriched library, clone:4631429H07, full insert sequence.
ACCESSION AK014540
VERSION AK014540.1 GI:12852460
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) 0 day neonate skin cDNA to mRNA, clone:lib:RIKEN full-length enriched mouse cDNA library
clone:4631429H07.

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 2880)
Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning
Methods in enzymology. 303, 19-44 (1999)

99279253

PUBMED

10349636

REFERENCE

2 (bases 1 to 2880)
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome research. 10 (10), 1617-1630 (2000)

11042159

PUBMED

20499374

REFERENCE

3 (bases 1 to 2880)
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Kono, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M.,
Sun, N., Ishii, Y., Nakamura, S., Hazama, M., Nishibe, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system - 384 format
Sequencing pipeline with 384 multicapillary sequencer
Genome research. 10 (11), 1757-1771 (2000)

REFERENCE

11076861

PUBMED

20530913

JOURNAL

MEDLINE

TITLE

AUTHORS

REFERENCE

PUBMED

TITLE

AUTHORS

REFERENCE

PUBMED

TITLE

AUTHORS

REFERENCE

PUBMED

TITLE

AUTHORS

REFERENCE

PUBMED

TITLE

AUTHORS

REFERENCE

PUBMED

TITLE

AUTHORS

REFERENCE

PUBMED

TITLE

AUTHORS

REFERENCE

PUBMED

TITLE

AUTHORS

REFERENCE

PUBMED

TITLE

AUTHORS

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gscc.riken.go.jp,
url:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)

Please visit our web site (http://genome.gsc.riken.go.jp/) for


```

180  rGlaAlaAlaLysValGluSerAlaAlaValThrGlnLysValG 197
    ||| ||||| ||||| ||||| ||||| ||||| |||||
173  .GAAGAAGCTAAATCAGATCAAGAAGGCA.....GCTAAGAAGGTAG 215
197  LuValLysPheSerLysAlaValGluLysLeuThrLysGluAspIleLys 213
    || ||||| ||||| ||||| ||||| ||||| |||||
216  AAAGAATAATG...AAAGAATCAAGAATAATCATTGAAGAGTAAAGATTAAAG 262
214  ValThrAsnLysAlaAsnAsnAspLysValLeuValLysGluValThrLe 230
    ||| ||||| ||||| ||||| ||||| ||||| |||||
263  GTTCACATCAAGACT.....GCTAAGAAGAGCTGCTAA 294
230  uSerGluAspLysArgSerAlaThrValGluLeuTyrSerAsnLeuAla 247
    :||| ||| :||| :||| :||| :||| :||| :|||
295  AGCAGAAAGACAGCTGAAAGCAATTAAGCAAGATTAAAG..... 337
247  LdLysGlnThrTyrThrValAspValAsnLysValGlyLysThrGluVal 263
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
338  .....GATGTTAAGAAGAAAGCACTAAGAGCTGCT 367
264  AlaValGlySerLeuGluAlaLysThrLeuGluMetAlaAspGlnThrVa 280
    :||| :||| :||| :||| :||| :||| :|||
368  CAACTCAAGAAGGTACAATCT..... 388
280  LvaAlaAspGluProThrAlaLeuGlnPheThrValLysAspGluAsnG 297
    ||| ||| ||| ||| ||| ||| |||
389  .....CAAGTTAAGAAGCAAGCAAGAAATG 413
297  LyrThrGluValValSerProGluGlyLeuGluPheValThrProAlaAla 313
    :||| :||| :||| :||| :||| :||| :|||
414  CTAAAGAAAGCTCTTTAAAGAAATCA...AAATTAAGAAATCAATTAGCT 460
314  GluLysIleAsnAlaLysGlyGluLeuThrLeuAlaLysGlyThrSerTh 330
    :||| ||||| ||||| ||||| ||||| ||||| |||||
461  AAGAAGATTAGTAAG.....ACTTCTAA 483
330  rThrValLysAlaValTyrLysLysAspGlyLysValAlaAlaGluSerL 347
    :||| :||| :||| :||| :||| :||| :|||
484  GAATGCTAAAGCTATTGCCAAGATCAGCTAAGAAAGCTGCTAAAAAG 533
347  YsgLysValLysValSerAlaGluGlyAlaAlaValAlaSerIleSerAsn 363
    ||| ||| ||||| ||||| ||||| ||||| |||||
534  CTGAAGAAAGAA.....GCAGAGAGAAAGCTATTACCAAGCAAGCAAGCT 577
364  TrpThrValAlaGluGlnAsnLysAlaAspPheThrSerLysAspPheLys 380
    :||| :||| :||| :||| :||| :||| :|||
578  TACACT.....AAGAAGATACCAAGAGATTGCTAAGAGATTAAGAA 621
380  sGlnAsnAsnLysValTyrGluGlyAsnAsnAlaTyrValGlnValGluL 397
    :||| :||| ||||| ||||| ||||| ||||| |||||
622  AAAACTTAACAAAGCTT..... 637
397  euLysAspGlnPheAsnAlaValThrThrGlyLysValGluTyrGluSer 413
    :||| :||| :||| :||| :||| :||| :|||
638  .....CAATCTAAG 646
414  LeuAsnThrGluValAlaValAlaAspLysAlaThrGlyLysValThrVa 430
    ||||| :||| ||||| ||||| ||||| ||||| |||||
647  TTGAAGAAAGCTTAAGAAAGTAAAGAA..... 676
430  LleuSerAlaGlyLysAlaLeuProValLysValThrValLysAspSerLysG 447
    :||| ||||| ||||| ||||| ||||| ||||| |||||
677  ....AATGATTAAGAGGCTTGTCTAAGATTAAGAAAGCAAGCAAGAAAG 722
447  LylLysAlaLeuValSerHisThrValGluLeuLeuAlaPheAlaGlnLys 463
    :||| :||| :||| :||| :||| :||| :|||
723  CACTTGTCTAAATTAATCT.....CAGAAA 745
464  AlMetLysAspIleLysLeuGlyLysThrAsnValAlaLeuSerThrLys 480
    :||| ||| :||| :||| :||| :||| :||| :|||
746  GTTAAGAAAGCTCAAGAAAGATTAAGAAAGCTATTGCTAAATTAAGCTAA 795

```

```

480  sAspValThrAspLysValLysAlaProValLeuAspGlnTyrGlyL 497
    ||| ||||| ||||| ||||| ||||| ||||| |||||
796  AAAGACACGAAGCTCTTCTAAATTCCTAGAAAGACAG..... 838
497  YsgLysPheThrAlaProValThrValLysValLeuAspLysAspGlyLys 513
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
839  .....AAAGCTATTGCTAAATTAAGCTAAG 862
514  GluLeuLysGluGlnLys 519
    :||| ||||| ||||| ||||| ||||| ||||| |||||
863  AAGCTAAAGAACCAAA 880
seq_name: gb_gss:BH135264
seq_documentation_block: 899 bp DNA GSS 07-AUG-2001
LOCUS BH135264 Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic DNA sequence.
ACCESSION BH135264
VERSION BH135264.1 GI:15094325
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 899)
AUTHORS Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
TITLES Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library (2001)
COMMENT Unpublished (2001)
CONTACT: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: entae@ligr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 15
High quality sequence stop: 801.
FEATURES
source
location/Qualifiers
1..899
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pROSL; Site 1. Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + 1 method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999)."
```

alignment_scores:

Quality:	141.00	Length:	309
Ratio:	0.839	Gaps:	15
Percent Similarity:	54.369	Percent Identity:	26.214

alignment_block:

US-09-754-947-1 x BH135264/rev ..

225 atlysgValValThrLeuSerGluAspLysArgSerAlaThrValGluLeu 241
225 atlysgValValThrLeuSerGluAspLysArgSerAlaThrValGluLeu 241
109 TCAGAAGA...TTGGAGAAACCAAGCTGAGGT... 140
242 TySerAsnLeuAlaLysGlnThrTyrThrValAspValAsnLysVal 256
141TTGGCGTTGGCAGACAGACAGCAGCAGCAGCAGTAAAGGC 181
258 LglLysThrGluValAlaValAlaGlySerLeuGluAlaLysThrIleGlu 275
182 GGAAAAAGTCGAA...GCTGTGCTATCAAGAGAACGTGAACAGAAAT 222
275 etAlaAspGlnThrValValAlaAspGluProThrAlaLeuGlnPheThr 291
229 CTCGCGCTGCCAAGAAATTTGAGAAACAAAAGCGCGCTGCCAAAGAA 275
292 ValLysAspGluAsnGlyThr.....GluValVal 301
279 TTGAGGACACAGAAATCCGTATTTGAAATTTGGAGCAGAAACAAAGTTGC 322
301 LserProGluGlyLe...GluPheValThrProAlaAlaGluLysIle 311
329 TGAGCGCCGAAAAATCAAGAGAGACAAACGCCCGCTACTGAGAAATTCG 372
317 snAlaLysGlyGluIleThrLeuAla....LysGlyThrSerThrThr 331
379 AAGTTAAAAAGCTGCTACTCTTATTTGAAACAGAAACAAAGCTGCTGCC 422
332 ValLysAlaValLysLysLys.....AspGlyLys 341
429 GCCACAGAAATTTGAGACACAAACACACATCTTTGAAATTTAGACAGAA 472
341 sValValAlaGluSerLysGluValLysValSerAlaGluGlyAlaVal 355
479 ACAAGTTGCACCGCAGAAAGAGATCGAAGACAAACAGCTGACGTGCCA 522
358 alAlaSerLieserAsnTPrThrValAla.....GluGlnAsnLys 372
529 AGCAATTTGAAAGTAAAAAGCGCAATTCCTTAAATCTAGAGCAAAAGCAA 572
372 AlaAspPheThrSerLysAspPheLysGlnAsnAsnLysValTyrGluGlu 388
579 GCT...GCTACTGCCAAGAAAT...GAAATTCAGAG... 611
388 yAspAsnAlaTyrValGlnValGluLeuLysAspGlnPheAsnAlaValT 401
612 ...TCTGCTGCTATCAAA...CAATTCAGAGAA... 630
405 hrThrGlyLysValGluTyrGluSerLeuAsnThrGluValAlaValAl 421
639AACCAAGCTGAGTTTGGCGCTTGGCAGACAGAACAGCACACCA 681
422 AspLysAlaThrGlyLysValThrValLeuSerAlaGlyLysAlaProVa 438
884 GCTTAAAAAGCGGCAAAAAGTCCAAAGT... 711
438 LlysValThrValLysAspSerLysGlyLysAlaLeuValSerHisThrV 455
711 ...GTTGCTATCAAGAGAGCTGAAGAACAAAAAGCTGCCACT...GCCA 755
455 alGluIleGluAlaPheAlaGlnLysAlaMetLysAspLleLysLeuGlu 472
754 AGCAAAATTTGAGCAACAAAAAGCGCGCTGCCAAGAAATTTGAGAACAA 801
472 LysThrAsnValAlaLeuSerThrLysAspValThrAspLeuLysValL 488
804 AAGACGCTGCTCTCCAAAAAGTAATTTGAAGGACAGCAAGATACCCTC... 845
488 sAlaProValLeuAspGluTyrGlyLysGluPheThrAlaProValThrV 505
846ATTTCGAATTTGGACAGCAAGAACAACTGCT... 877
505 alLysValLeuAspLysAspGlyLysGluLeuLysGluGlnLysLeuGlu 521

131 IuThrLeuAsnTrpGlySerGluValAlaAsnIle..LeuValGluLeuG1 147
||||| : : : : :
409 AAGCACTACACACACGACGATCCGAGCCTCATCTGCTGTCG..... 450
147 yIleSerValGlyThrGlyAspGlnTrp.....GluProL 159
: : : : :
451 .CTCAGTCTGCACACTGAGCGTCTGCTACCAAGTGTGACGAGAG 499
ySlyThrValThrLysAlaGluAlaValGlnPheIleAlaLysThrAsp 175
||| : : : : :
500 TCAAGTACTGTAAGCTCAACCGCATTTGGGCCAAGTGTGATTATGTTGA 549
176 LysGlnPheGlyThrGluAlaAlaLysValGluSerAlaLysAlaValThr 192
||| : : : : :
550 AAACAGCTGCGCTGAAGAACTTCAAAACACACAGAGAAA.....AA 590
192 rThcGlnLysValGluValLysPheSerLysAlaValGluLysLeuThrL 209
: : : : :
591 TAAATGACACACTGAGCTCGAA...AATTAATAATTTGAGAAAGAGACTA 637
209 ySGluAspIleLysValThrAsnLysAlaAsnAsnAspLysValLeuVal 225
||| : : : : :
638 AAATATGAA...CAAGAGACAGAGAAATCGGAAAC.....CTGGCT 675
226 LysGluValThrLeuSerGluAspLysArgSerAlaThrValGluLeuTy 242
||| : : : : :
676 AAAGAACTATTTCCATGACTCTGCTTCCAGATTAACCTGTGAAGAGACT 725
242 rSerAsnLeuAlaLysGlnThrTrpThrValAspValAsnLysValG 259
||||| : : : : :
726 CAGTAAATTTGGGGAATCTGTTTCTTCAATGCACTTAATGCAAGACTTGT 775
259 LyLysThrGluValAlaValGlySerLeuGluAlaLysThrIleGluMet 275
: : : : :
776 CACAAAGCCGACGTGCTTACAGACTACTA.....AAGAGTGAAGATG 819
276 AlaAspGlnThrValAlaAlaAspGluProThrAlaLeuGlnPheThrVa 292
: : : : :
820 TCTTGA...ACGATTGTAAATAAGAGCCACCTCATCTGGCCTAACA.. 864
292 LysAspGluAsnGlyThrGluValValSerProGluGlyIleGluPheV 309
864 864
309 aIThrProAlaIaGluLysIleAsnAlaLysGlyGluIleThrLeuAla 325
||| : : : : :
865 .GAACCTTTAGAACTAAACCTCGAGCCTCCAGCTCTTACTTTTACCC 912
326 LysGlyThrSerThrThrValLysAlaValTyrlsLysAspGlyLysVa 342
913 ATG.....AGCCAGTT 923
342 lValAlaGluSerLysGluValLysValSerAlaGluGlyAlaAlaVala 359
: : : : :
924 TCTCAGTGAAGTGCACAGAACAAA..... 948
359 lAserIleSerAsnTrpThrValAlaGluGlnAsnLysAlaAspPheThr 375
949AAGCGAGTTGTGACA 963
376 SerLysAsp...PheLysGlnAsnAsnLysValTyrlGluGlyAspAsnAl 391
964 CCTTAAAGAGCTTTTCTTCAAGCTGTGTAATAAAACCAACACGTTTAAAG 1013
391 aTyrlValGluValGluLeuLysAspGlnPheAsnAlaValThrThrGlyL 408
: : : : :
1014 GTACACACAAACAGACAGCAGAGCTCTTCCTACTACTGATGGGA 1063
408 ySValGluTyrlGluSerLeuAsnThrGluValAlaValAspLysAla 424
||| : : : : :
1064 TGACAGCGCAAAACACCAAAAGACTGAGTAAAGAAATTTCTT...AAGCA 1110

425 ThrGlyLysValThrValLeuSerAlaGlyLysAlaProValLysValThr 441
||||| : : : : :
1111 TTGGTAAATTTCTACT.....GAAAAATTGCATGAAGAAGTAAAAAATA 1154
441 rValLysAsp...SerLysGlyLysAlaLeuValSerIleThrValGluI 457
||||| : : : : :
1155 AGTTAAAGATTATGAAAAAGAAAAAGCAATCCCGAGTTTGTGGACCCCA 1204
457 lGluAlaPheAlaGlnLysAlaMetLysAspIleLysLeuGluLys... 472
||| : : : : :
1205 TC.....TTTGGTGGCAGCTGACTAGTACGATCATGTGTGATAAATGC 1248
473 ThrAsnValAlaLeuSerThrLysAspValThrAspLeuLysValLysAl 489
1249 AGAGCTCTCTCTTATGTCATGAATCGTTCTTATTTGTCTCTT..... 1293
489 aProValLeu...AspGlnTyrlGlyGluPheThrAlaProValThrV 505
||||| : : : : :
1294 .CCAGTTTATGATGATCAGAGTGTAAAGAAAAGTAAATGATAAAAAATG 1342
505 aLlysVal...LeuAspLysAspGlyLysGluLeuLysGluGlnLysLeu 520
||||| : : : : :
1343 TGAATAATGACATGAGAGAGAGATTAAGACAGCTGAGAGAGAAA... 1389
521 GluAlaLysTyrlValAsnArgGluLeuValLeuAsnAlaAlaGlyGlnG1 537
: : : : :
1390 GATGACAGCTACATGAATAATCAAGAGCGATCTTCGTCAGGG..... 1431
537 uAlaGlyAsnTyrlThrValValLeuThrAlaLysSerGlyGluLysGluA 554
||||| : : : : :
1432ACAAGCAAGCACTTACAGAAAAGG 1456
554 lAlysAlaThrLeuAlaLeuGluLeuLysAlaProGlyAlaPheSerLys 570
||||| : : : : :
1457 CAAGAAACAGAGCCAAAAGACAGCCCAAGAACCAAGAGCAACAAAA 1506
571 PheGluValArgGlyLeuAsp...ThrGluLeu.....AspLysTyrlVa 584
: : : : :
1507 ATTCAAGAAATTTCTTCACTTCAATGAGCTCTGCCACTGACTAC.. 1554
584 lThrGluGluAsnGlnLysAsnAlaMetThrValSerValLeuProVala 601
||||| : : : : :
1555 .ACGGAAACAAATGACCTGAAGCTGACACA..... 1584
601 sPalaAsnGlyLeuValLeuLysGlyAlaGluAlaAlaGluLeuLysVal 617
||| : : : : :
1585GCACCTGGCGGAGAGTGAAGTGAAGTGA 1608
618 ThrThrThrAsnLysGluGlyLysGluValAlaAspAlaThrAspAlaGlnVa 634
||| : : : : :
1609 GATACCGACTCCACCCATGCTTCCAAAGAGAGGCCACACAGATAGAGCT 1658
634 lThrValGlnAsnAsnSerValIleThrValGlyGlnGlyLysAlaG 651
: : : : :
1659 GTCTGTTAACCAAGAGAT..... 1677
651 LyGluThrTyrlLysValThrValValLeuAspGlyLysLeuIleThrThr 667
1678TTGATGGCCCAAGAGAGCATGATA 1701
668 HisSerPheLysValValAspThrAlaProThrAlaLysGlyLeuAlaVa 684
: : : : :
1702 GAAAGGACACCTGATGTGCAGAAAGCCCAAGAGACCTTAGAGAAAG 1751
684 lGluPheThrSerThrSerLeuLysGluValAlaProAsnAlaAspLeuL 701
||| : : : : :
1752 TGCTAACACCCAGAGTATCTGGGATTTGACGCT..... 1788
701 ySAlaAlaLeu..LeuAsnIleLeuSerValAspGlyValProAla 715
||||| : : : : :
1789GCTCCTGAATGTCCTAGGATTTCAATGATGCTTCC 1825

seq_name: gb_hlc:BC008524

seq_documentation_block:
LOCUS BC008524 4379 bp mRNA HTC 12-JUL-2001
DEFINITION Mus musculus, similar to lipoprotein-binding protein, clone
IMAGE:3601469, mRNA.
ACCESSION BC008524
KEYWORDS BC008524.1 GI:14713711
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 4379)
REFERENCE 1 (bases 1 to 4379)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (25-MAY-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
REMARK Contact: MGC help desk
COMMENT Email: cgapsb-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) Genome
DNA Sequencing by: Sequencing Group at the Stanford Human
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mdickpaxl@stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAP Plate: 11 Row: f Column: 22
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, genomeScan gene prediction, similarity but not identity
to protein
This clone has the following problem: frame shifted.
Location/Qualifiers
FEATURES
source 1. 4379
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3601469"
/tissue_type="Mammary tumor. Metallothionien-TGF alpha
model. 10 month old virgin mouse. Taken by biopsy."
/clone_lib="NCI CGAP_Mam1"
/lab_host="DH10B"
/note="Vector: PCMV-SPORT6"
BASE COUNT 1306 a 1063 c 1113 g 897 t
ORIGIN
alignment_scores:
Quality: 137.00 Length: 969
Ratio: 0.314 Gaps: 49
Percent Similarity: 45.098 Percent Identity: 18.679
alignment_block:
US-09-754-947-1 x BC008524 ..
Align seg 1/1 to: BC008524 from: 1 to: 4379
45 GUAUAAAThmetMetAlaGlnIleLeuasnLeuproIleasplysas 61
:::|||||::: ||| |||
208 AAAGTTCACCTTAATTCAGAAAGAGAAATACCT..... 246
61 palalysProserPhealIaasPserInglyGlnTpyrThrPropei 78
|||::: |||::: |||
247CCAACTCAAGATGCC.....TTCC 268
78 leaIaIaValaGluLysAlaGlyValIleLysGlyThrGlyAsnGlyPhe 94

|||||:::
269 CTCACCTTCTGAGAAAGCAGCTTGTGGAG.....ATGCCCAA 309
95 GluProasnGly.....LysIleasParyalsermetIase 107
||||| ||| ||| ||| |||
310 GAACCTGCCGAGCGCTGAGATCAACAGATC...CGGCCATCAAGGCTC 346
107 rLeuLeuValaGluAlaLysLysLeuaspThrLysValaAsnGlyThrPro 144
|||||:::
357 TGTTCATCCTCAGGTGTTTCATGTACCTCGAA.....G 391
124 laThrLysPheLysAspLeuGluThrLeuasnTrpGlyLysGluLysAla 140
|||||:::
392 AGAGAAATATATAGGACATGACACAGTTGGAGAGGT...GAACAAGA 438
141 AsnIleLeuValaGluLeu..... 146
439 AAATATCTGCTTGATCATGACAGAGAGACTGTGCTCACCCTGCACTCTC 488
147GlyIleSerValaGlyThrGlyAspGln..... 155
489 TGTGGCTAAGATCAAGATCAAGTCTTCATCATGATGCTCAGGAAGCTGGATG 538
156TrpGluProLysLysThrVal.....ThrLysAla 165
539 CCGTCATGAAGACCCGCAAGACATTTGCTAGACTACAGACACAGGCC 588
166 GluAlaIaGlnPheIleAlaLysThrAspLysGlnPhe.....GlyThr 180
589 TCAGCACTGTCCCATTCCTTAAAGACACCATCGCTTTGTTATGGCAA 638
180 rGluAlaIaLysValaGluSerAlaLysAlaValThrGluLysValaG 197
639 AAACGAGAGAAACTGCAGACCTAGAACTAAACATCAACCAAAATCC 688
197 luValLys.....Phe 200
689 AGATCCACGCCGATGACCCAGTATATGATCAAGATCACTGGCACC 738
201 SerLysAlaValaGluLysLeuThrLysGluAspIleLysValThrAsn 217
739 AAGAGAGCATCGAGAAAGCTCGGCATGAGTCTCTATATCT..... 783
217 salAAsnAsnAspLysValaLysGluValThrLeuSerGluAspL 234
|||||::: ||| ||| ||| |||
784 .CGTAGACAGAGACAGCGCTGCTGAGAGAGCTAAGAGTGAAGGCTT 832
234 ysarGserAlaThrValaGluLeuThrSerAsnLeuAla..... 247
833 TCCACCCCTTCATCTGCGCGCTGACACAGACTTGGGTGAATCATG 882
248 LysGlnThrTyrThr.....ValAspValAsnLysValaGlyLysTh 261
883 CAGGAGACAGGAGCGCAATCAATCCCTCCACCCAGTGTAAACGTAC 932
261 rGluValaIaL...ValGlySerLeuGluAlaLysThrIleGluMetAla 277
|||||:::
933 AGAATCGTCTTCACTGAGAGAGAGAGAGAGCTGCGCTGAGCTGCTC 982
277 spGlnThrValaIaAspGlu.....ProThrAlaLeuGln 289
983 GCATTAAAGAGATTATAGGAGAAAGAGAGAGAGACTACACCATTTGCG 1032
290 PheThrValLysAspGluAsnGlyThrGluValaValSerProGluGlyI 306
|||||:::
1033 GTAGAGTGAAGAGTCTCAGACCAATATGTATTGGCCAAAGGCC... 1080
306 eGluPheValThrProAlaIaGluLysIleAsnAlaLysGlyGluIle 323
::: ::: |||::: |||::: |||
1081 .AATTCACTGCAAGAAATCTCGAAAGAACTGAGTTCTGTGTTGAGATCC 1129
323 hrLeuAlaLysGlyThrSerThrVal.....LysAlaValTyrLys 337
::: ::: ||| ||| ||| |||

```

1130 CACCCTCGAGCAGCATCTCTGAGACTGTGATCTTATGAGCGCAACTGAA 1179
338 LysAspGlyLysValAlaGluSerLysGluValLysValSerAlaGlu 354
1180 AAGTTGGGACAGCGCTTACT.....GAGCTTATGCCAAGGCCAA 1220
354 uGlyAlaAlaValAlaSerIleSer.....AsnTrpThr..... 365
1221 CAGTTTACACTCTCTCAGTCTGTGCCCCCTTCGCTTCATCGTTTCA 1270
366 .....ValAlaGluLysAsnLysAlaAspPheThrSerLysAspPhe 379
1271 TCATTGTGTAAGAAAGGCGAAGCCTGCGCAAAATCTCAGCAAAATGCCA 1320
380 LysGlnAsnAsnLysValTyrGluLysAsn..... 390
1321 AAGTTTCACATTGAGTTCCAGAAAGTGAAGACAGATCACCCTGAGGG 1370
391 .....AlaTyrValGlnValGlu.....LeuL 398
1371 CCCCACAGAGCATGTGAATGTAGCCCAAGAACAGATTGAAGAAATGCTCA 1420
398 LysAspGlnPheAsnAlaValThrThrGlyLysValGluTyrGluSerLeu 414
1421 AAGACCTGATTAAAT.....AGGATGGACTATGTAGAGATC 1455
415 AsnThrGluValAlaValAlaAspLysAlaThrGlyLysValThrVal.. 430
1456 AACATTGACCTAAATTCACAGACACTCATTTGAAAGACGCGAGCCAA 1505
431 LeuSerIleGlyLysAlaProValLysValThrValLys.....A 444
1506 CATCAACAGAAATCAAAAGACAGTCAAAAGTCTGTGCGCATCCCCCTG 1555
444 sPseLysGlyLysAlaLeuValSerIleThrValGluIleGluAlaPhe 460
1556 ACAGTGAA.....AAGAGTAACTGATCCGATTCGATGAGGGAAC 1593
461 AlaGln..... 462
1594 CCACAGGGGTTCAGACAGGCCAGGGAGCTGCTGAGCTTGTCTCG 1643
463 .....LysAlaMetLysAspIleLysLeuGlu..... 471
1644 CATGAAAATAGCGCACCAAGATCTAATCATGAGCAAAAGATTTCATC 1693
471 ..... 471
1694 GCACAATCATTTGGCAGAAAGTGAACGATTCGAGAAATTCGTGACAA 1743
471 ..... 471
1744 TTCCAGAGCTTATCATCACTTCCAGACCCACACAAAAGTGAATAT 1793
472 .....LysThrAsnValAlaLeuSerThr..... 479
1794 TGTCAACTCAGAGCCCCCAAGATGAGGTGAGAGATGCTACATAAGTATA 1843
480 .....LysAspValThrAspLeu.....LysValLysAla 489
1844 TGCAGAAAGATGTGCGACAGCTGTGGAATAATACCTATTCAATTCTGTA 1893
490 ProValLeuAspGlnTyrGlyLysGluPheThrAlaProValThrValLys 506
1894 CCATCTTCAAAACAGTTTCACAG.....AA 1919
506 sValLeuAspLysAspGlyLysGluLeuLysGluLysLeuAlaL 523
1920 TATCATCGCGGAAAGAGCGCAACATTAAAGATCCGTGAAGAGAGC. 1968
523 ySTyrfValAsnArgLysLeuValLeuAsnAlaAlaGlyGlnGluAlaGly 539
1969 .....AMCACCAGATTGACTCTCTGCA.....GAGAACAGCAAC 2004

```

```

540 AsnTyrThrValValLeuThrAlaLysSerGlyLysGluAlaLys.. 555
2005 TCTGAGACCATATTATTCACCTGCGCAAGCAGCAACTGTGAAGCTGCCG 2054
556 .....AlaThrLeuAla...LeuG 561
2055 CAGCGGATTCGTCTGTATCCAGAAAGATCTGGCCACATAGCTGAAGTGG 2104
561 LysLeuLysAlaPro..... 565
2105 AGGTCTCAATTCTGCAAACTACACAACTCTCTCATTTGGCACAAAGGC 2154
566 .....GlyAlaPheSerLysPhe 571
2155 CGTCTGATCCGCTCCATCATGAGAGAAATGTCGCTCCATTCATCTACT 2204
571 eGluVal.....ArgGlyLeuAspT 578
2205 TCTGTGGAAGGTTCTGGAAGCGACACTGTGTCTCATCAGGGCTCTTCT 2254
578 hrgLysLeuAspLys.....TyrValThrGluLysAsn 588
2255 CAGATTGTAAGAAAGCCAGACAGCCTCCTGCACTGGCTGAGAGAGAG 2304
589 GlnLysAsnAlaMetThrThrValSerVal.....LeuProValAspAlaAs 603
2305 CAACCAAGAGATTCTACTGTGACATCCGAGCCAGACAGAAATCCACAA 2354
603 nGlyLeuValLeuLysGlyAlaGluAlaAlaGluLeuLysValThrThr 620
2355 ATTCCTCATTTGGCAAGGGGA..... 2376
620 hTrAsnLysGluGlyLysGluValAspAlaThrAspAlaGlnValThrVal 636
2377 ..GGTAAATCCGAAAGGTGCGTGCACAGTACTGAGGCCCGCATCATATT 2424
637 .....GlnAsnAsnSerValIleThr...ValGlyGlnGlu 647
2425 CCGGCTGCTGAAGATGAAGACACAGACCTGATCAATCATAGGAAGGA 2474
647 yAlaLysAlaGlyLysLysThrTyrLys..... 655
2475 GGATGCTGTCGCTGAGGCCCAAAAGAGCTGGAAACATTGATCCAGAAC 2524
656 ..ValThrValValLeuAspGlyLysLeuIleThrThr.....HisSer 669
2525 TGGAGATGTGCTTGAAGACATATGCTGTGGACCCCAAAACACACCGC 2574
670 PheLysValValAspThrAlaProThrAlaLysGlyLeuAlaValGluP 686
2575 CATTTGTCTCATCCGAAAGAGCAAGTTTACGAGAAATTCCTGAAAGGTA 2624
686 eThrSerThrSerLeuLysGluValAlaProAsnAlaAspLeuLysAla 703
2625 TGCGGGGGTATGCTCAGC.....TTCCACGCTGTGCAAGCAGAGGTG 2668
703 lAlaLeuAsnIleLeuSerValAspGlyValProAlaThrThrAlaLys 719
2669 ATAAAGTCACTCTCAAGGTCGCCAAGAGCTGTGAGGAGCGCAAGCAAAA 2718
720 AlaThrAlaSerAsnValGluPheValSerAlaAspThrAsnValAlaL 736
2719 CGCATTCAGAAATCATTTGAGATCTGAAGCTCAATGACAGAGTAAGT 2768
736 a.....GluAsnGlyThrValGlyAlaLysGlyLysAlaT 747
2769 TGCTATATCCCGAAGATTTCACCGATGCTGATGCGCCCAAGGTTCA. 2817
747 hrSerIleTyrValLysAsnLeuThrValValLysAspGlyLysGluGln 763
2818 .....AGATTCAG 2826

```


[illegible]

VERSION AK019969.1 GI:12860394
KEYWORDS CAP trapped.
SOURCE Mus musculus (strain:C57BL/6J) 8 days embryo cDNA to mRNA,
clone:5730504C04.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2413)
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Methods in enzymology. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
2 (bases 1 to 2413)
REFERENCE Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
AUTHORS Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
JOURNAL prepare full-length cDNA libraries for rapid discovery of new genes
MEDLINE 20499374
PUBMED 11042159
3 (bases 1 to 2413)
REFERENCE Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
AUTHORS Konno,H., Akiyama,J., Nishi,K., Kitsuana,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanabe,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
JOURNAL sequencing pipeline with 384 multicapillary sequencer
MEDLINE 20530913
PUBMED 11076861
4 (bases 1 to 2413)
REFERENCE The RIKEN Genome Exploration Research Group Phase II Team and the
AUTHORS FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
MEDLINE 5 (bases 1 to 2413)
PUBMED
AUTHORS Adachi,J., Aizawa,K., Akeshira,S., Akimura,T., Aono,H., Arai,A.,
Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
Hangaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hirooka,T., Horii,F.,
Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
Kanagawa 230-0045, Japan (E-mail:genome-resgsc.riken.go.jp,
URL:http://genome-gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome-gsc.riken.go.jp/) for
further details.
FEATURES
source Location/Qualifiers
1. 2413
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="MGI:1911416"
/db_xref="MGI:1924829"
/clone="5730504C04"
/clone.lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="8 days embryo"
BASE COUNT 783 a 500 c 798 g 332 t

ORIGIN
alignment_scores:
Quality: 135.00 Length: 515
Ratio: 0.511 Gaps: 24
Percent Similarity: 51.262 Percent Identity: 21.942
alignment_block:
US-09-754-947-1 x AK019969 ..
Align seg 1/1 to: AK019969 from: 1 to: 2413
151 G1YThrGlyAspGlnTyrGluProLysLysThrValThrIysAlaGluAl 167
||| ||| :||| ||| :|||
173 GGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 222
||| ||| :||| ||| :|||
167 aAla.GlnPheIleAlaLysThrAspLysGlnPheGlyThrGluAlaAla 183
||| ||| :||| ||| :|||
223 TGACGGCGACATTCAGACCTAGAGACAACTGATGAGAGAG..... 266
184 LysValGlnSerAlaLysAlaValThrThrGlnLysValGluValLysph 200
||| ||| :||| ||| :|||
267GAGGGGCCCG.....CAAAAGCTGACGTCGAGAA 298
200 eSerLysAlaValGluLysLeuThrLysGluAspIleLysValThrAsnL 217
||| ||| :||| ||| :|||
299 GGTACAGACAGAGCTTAAATTCAGAGAGATGAGAGAGAGAGAGAGAG 348
||| ||| :||| ||| :|||
217 yAlaAsnAsnAspLysValLeuValLysGluValThrLeuSerGluAsp 233
:||||| :||| ||| |||
349 TCGAAGACCAAGAAATTCAAATTATCAAGAGAGAGAGAGAGAGAGAC 398
||| ||| :||| ||| :|||
234 LysArgSerAlaThrValGluLeuTyrSerAsnLeuAlaLysGlnTh 250
||| ||| :||| ||| :|||
399 CGA.....ATGCTGAGTCTCTCTCTCAGCTGCTGAGAGAGAGA 439
||| ||| :||| ||| :|||
250 rTyrThrValAspValAsnLysVal...GlyLysThrGluValAlaValG 266
||| ||| :||| ||| :|||
440 AAGAGCAAAACCTTGCCCAAAATCAGAGAAATGACAGAGAGAGAGAG 489
||| ||| :||| ||| :|||
266 LysLeuGluAlaLys.....ThrIle 273
||| ||| :||| ||| :|||
490 CGGACTTGAAGAGAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 539
||| ||| :||| ||| :|||
274 GluMetAla.....AspGlnThrValAlaAlaAspG 284
||| ||| :||| ||| :|||
540 GAAAGGCCAAAGCAGAGCTGATGGGAGAGAGAGAGAGAGAGAGAGAG 589
||| ||| :||| ||| :|||
284 uProThrAlaLeuGlnPheThrValLysAspGluAsnGlyThrGluValY 301
:||||| ||| ||| |||
590 GATGCTGAGCTGACGAGCAGACAGGTC...GATGAGCTCAAGTCACAGTGA 636
||| ||| :||| ||| :|||
301 aLysrProGluGlyIleGluPheValThrProAlaAlaGluLysIleAsn 317
||| ||| :||| ||| :|||
637 CCAAGAGAGAG.....GAGGAGCTTCAG 659
||| ||| :||| ||| :|||
318 AlAlysLysGluIleThrLeuAlaLysGlyThrSerThrValLysAl 334
||| ||| :||| ||| :|||
660 GCG.....GGCTGGCCAGAGAGAGATGATAGACA..... 689
||| ||| :||| ||| :|||
334 aValTyrLysLysAspGlyLysValValAlaGluSerLysGluValLysV 351
:||||| :||| ||| :|||
690 .CTGCACACAGAAATATGCACCTTAAGATTGACAGG.....GAGC 726
||| ||| :||| ||| :|||
351 aLysrAlaGluGlyAlaAlaValAlaLysrIleSerAsnTyrThrValAla 367
||| ||| :||| ||| :|||
727 TGCAGGCCCAATTCGACAGAGCTCAGAGACCTTGTGCTGAGAAAGCT 776
||| ||| :||| ||| :|||
368 G1G1G1AsnLysAlaAspPheThrSerLysAspPheLysGlnAsnLys 384
:||||| :||| ||| :|||
777 TCAAGGACACAGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 826

seq_name: gb_hic:AK019450

seq_documentation_block: 2633 bp mRNA HMC 05-JUL-2001

LOCUS AK019450 Mus musculus 14, 17 days embryo head cDNA, RIKEN full-length DEFINITION enriched library, clone:323040Z02, full insert sequence.

ACCESSION AK019450

VERSION AK019450.1 GI:12859660

KEYWORDS CAP trapper.

SOURCE	ORGANISM	REFERENCE AUTHORS TITLE JOURNAL MEDLINE PUBLISHED REFERENCE AUTHORS	TITLE JOURNAL MEDLINE PUBLISHED REFERENCE AUTHORS	
Mus musculus (strain:C57BL/6J) 14, 17 days embryo head cDNA to mRNA, cloneLib.RIKEN full-length enriched mouse cDNA library clone:3230402H02.	Mus musculus	Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	TITLE JOURNAL MEDLINE PUBLISHED REFERENCE AUTHORS	
1 (bases 1 to 2633)	Carninci,P. and Hayashizaki,Y.	High-efficiency full-length cDNA cloning Methods in enzymology. 303, 19-44 (1999) 99279253	2 (bases 1 to 2633)	Shibata,K., Carninci,P., Itoh,M., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
10349636	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome research. 10 (10), 1617-1630 (2000)		JOURNAL MEDLINE PUBLISHED REFERENCE	3 (bases 1 to 2633)
20499374	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasakhi,N., Carninci,P., Kono,H., Akiyama,T., Nishi,K., Kitsumai,T., Tashtro,H., Itoh,M., Suni.N., Ishii,Y., Nakamura.S., Hazama.M., Nishine.T., Harada.A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaka,E., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Matsuhiki,M., Yoneda.Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuda.S., Kawal,J., Okazaki,Y., Muramatsu,M., Inoue.Y., Kita,A. and Hayashizaki,Y.	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome research. 10 (11), 1757-1771 (2000)	4 (bases 1 to 2633)	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001) 5 (bases 1 to 2633)	Adechi,J., Aizawa,K., Akahira,S., Akimura,T., Anono.H., Arakawa,T., Carninci,P., Fukuda,S., Fukushima,Y., Furuno.M., Hanagaki,T., Hara,A., Hayatsu.N., Hiramoto,K., Hirokawa,T., Horl,F., Imotani,K., Ishii,Y., Itoh,M., Izawa.M., Kato,H., Kawasaki.J., Koima,Y., Kono,H., Kouda.M., Koya.S., Kurihara.C., Matsuyama,T., Miyazaki.A., Nishi.K., Nomura,K., Numasaki,R., Ohno.M., Okazaki,Y., Okido,T., Owa,C., Saito.H., Saito,R., Sakai,C., Sakhi,K., Sano.H., Sasaki,D., Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T., Sugabe,Y., Suzuki,H., Tagami,M., Takahashi.F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshiuo.M., Muramatsu,M. and Hayashizaki,Y.	Direct Submission Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Riken Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gscc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)	Please visit our web site (http://genome.gsc.riken.go.jp/) for further details. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGACAGATCCAGATCCAGATCTCTTTTTCCTTTTTTNN 3'], cDNA was prepared by using reverse thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 60.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTTCATAATTAAATTAATCATCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after	COMMENT

bulk excision from lambda F1C I. Cloning sites, 5' end: SalI; 3'
end: BamHI. Host: DH10B.
Location/Qualifiers
source
1. 2633
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="MGI:MGI:1911407"
/db_xref="MGI:MGI:1924822"
/clone="3230402H02"
/isuse.type="head"
/clone_lib="Riken full-length enriched mouse cDNA library"
/dev_stage="14, 17 days embryo"
172..2343
/note="putative"
/protein_id="BAB31726.1"
/db_xref="GI:12859661"
/translation="MGFCGSSRNQOMLSTEDGKFLIKSGCLGIGISNSRPPQPAV
ATPCADAPRMCHAGEFLEVENTSLFKQKHKKVVKISKYLDWMKLVINKERL
ANESLCLKQAGLGEVSVTRARNSASQIATANTVPSPSHSNTETFLSTETL
RSTAEFLGTAETLRNTAKTGLSTAEFLTAKTGLSTAEFLTAKTGLSTAKTGLS
TAKTGLSTSEAYSSOSSKRGPLHLHAGATDESMSPLTTPPSSITTEGVAEOVCN
FTLESRVSSLSASIQMRTPFAPSPCNLSLYSDTSGPMCHPRLIDNFTYGGNKDLO
ACTVYVRRIVSLDGEESTVLTQDPLPARPEYNSKRTSTLQVWMTSSKVSWE
VQLFBNDRKIOBOVOESTVTSOTTFLELNENSTKVAITVSGEKRSFPYINIST
VSPVVDLGIISPNLSLISWSRSGNVEOYRLVLDKAIYQDNVDRDYSVAHE
LTPGHLYNLTIVTMASGLONSRLKVRTPMEVSNLKVNDRLSLANKMOKPRDY
DESITLSHQGTIKESKTLAPVTEFOFKDLVGRLOYTICISGELSASRSAGRT
VPEKVRNLVSYNDIMKSFYVMTWTPAGDMWRYRLVFNESLVLNTITVKEETHAL
DGLFLIPGROYELEVESGNLRNSRCGRGRKSNLIORT"

BASE COUNT 718 a 665 c 643 g 607 t
ORIGIN

Alignment_scores:
Quality: 134.50 Length: 706
Ratio: 0.388 Gaps: 32
Percent Similarity: 49.150 Percent Identity: 18.839

Alignment_block:
US-09-754-947-1 x AK019450 ..

Align seg 1/1 to: AK019450 from: 1 to: 2633

```
146 leuGlIyIleSerValGlyThrGlyAspIntrpGluProLysLysThrVa 162
|||||
259 CTGGGTATCTCCAACTCTCAAGGGCCCTTCCACACGACGATTGGTAC 308
|||||
162 lThrLysAlaGluAlaAlaGlnPheIleAlaLysThrAspLysGlnPheG 179
|||||
309 TCCTGTGCCCAGGACCCCGATGACCTGCATCGCAGGAAGGTTTC. 357
|||||
179 lYThrGluAlaAlaLysValGluSerAla.....LysAlaVal 191
|||||
358 .....CTTGAGGTGGAAAAATACCTCAGCTCTTCTTCAGAAACAA 396
|||||
192 ThrThrGlnLysValGluValLysPheSerLysAlaValGluLysLeuTh 208
|||||
397 AACCATTAAGTAGTGTGCAAGAGATCATGTAATACCTCGATACCTGGAT 446
|||||
208 rLysGluAspIleLysValThrAsnLysAlaAsnAsnAspLysValLeuV 225
|||||
447 GAACTTAGTATTCAACAAGAGGAGAGCTGCCAATGAAGACCTCGTGC 496
|||||
225 aLlysgLysValThrLeuSerGlnLysAspLysGlnValGluLeu 241
|||||
497 TGAAGCAAGCTGCGCTGGTACAGAA.....GTCTCGCTCGGACT 537
|||||
242 TyrSerAsnLeuAlaAlaLysGlnThrTyrThrValAspValAsnLysVa 258
|||||
538 GCTGAAAACTAGCTGCTTCCACAGATTGCCACTACT...GCTAACACCGT 584
```

```
258 l.....GlyLysThrGluValAlaValGlyLys 267
|
585 |||||
585 CCATATAGCCCAAGCCACATTACCAATACACAGAGACCTCCCGGGA 634
|||||
267 erLeuGluAlaLysThrIleGluMetAlaAspGlnThrValAlaAlaAsp 283
|||||
635 GC...ACTACAGAGACCTCCAGAGACCTCAGAGACCTCGGGAGACT 681
|||||
284 GluProThrAlaLeuGlnPheThrValLysAspLysGlnGlyThrGluVa 300
|||||
682 GCAGAGACC...CTCAGAGACACTGCAGAGACCTCGGGAGACT.... 723
|||||
300 lValSerProGluGlyIleGluPheValThrProAlaAlaGluLysIleA 317
|||||
724 .....GCAGAGACCTCAGAGACCTGCAGAGACCTCGGGAGACT 757
|||||
317 snAlaLysGlyGluIle.....ThrLeuAlaLysGlyThrSerThrTh 331
|||||
758 GGACACTGCAGAGACCTCAGAGACCTGCAGAGACCTCGGGAGACT 807
|||||
332 ValLysAlaValTyrLysLysAspGlyLysValAlaGluSerLysG 348
|||||
808 GCAAGACCCCTC..... 819
|||||
348 uValLysValSerAlaGluGlyAlaAlaAlaSerIleSerAsnTrp 365
|||||
820 .....GGGACACTGCAGAGACCTCGGGAGC.... 846
|||||
365 hrValAlaGluGlnAsnLysAlaAspPheThrSerLysAspPheLysGln 381
|||||
847 .....ACTTCAGAGCCTACAGTCAG 867
|||||
382 AsnAsnLysValTyrGluGlyAspAsnAlaTyrValGlnValGluLeu 398
|||||
868 AGCAGACAGCAAGAGAGCTCCCCATCTGCACACAGCGCTGAGCTACAGA 917
|||||
398 sAspGlnPheAsnAlaValThrThrGlyLysValGluTyrGluSerLeuA 415
|||||
918 CGAATCTCTGAGCCCGCTGACTACT....CCACCTTCTCCAGCATCA 961
|||||
415 snThrGluValAlaValAlaLysAspLysAlaThrGlyLysValThrValLeu 431
|||||
962 CCACAGAGACTGCGAGTGGCAGAGCACTGAATGTAATCTACACTGTG 1011
|||||
432 SerAlaGlyLysAlaProValLysValThrValLys..... 443
|||||
1012 GAGTCCAGAGCTCTGACTGTGTCAGCTATCCAGGAGAGACTTTCG 1061
|||||
444 .....AspSerLysGlyLysA 449
|||||
1062 GTCACCTGTAACCTTAGCTCATCTACACAGAGTGAATCCTCGGG...C 1108
|||||
449 lAlaValSerHisThrValGluIleGluAlaPheAlaGlnLysAla... 464
|||||
1109 CCAATGCTGCGATCCATTCGATGATACACAATTTACTTACCGATGTAC 1158
|||||
465 MetLysAspIleLysLeu..... 470
|||||
1159 CCCAAGATTTTACAGAGAGGACCGCTATTAATCTCAGATGTTTCTCT 1208
|||||
471 .....GluLysThrAsnValAlaLeuSerThrLysAspValThrAspL 485
|||||
1209 GAGTGAAGAGAGAGCACTGCTGCTTACAGAGATCCGCTTG.... 1251
|||||
485 euLysValLysAlaProValLeuAspGlnIlyrGlyLysGluPheThrAla 501
|||||
1252 .....CTCTGCGCAGGTTTGAAGTCAATTCGGGAAAAACAGCA 1290
|||||
502 ProValThr..... 504
|||||
1291 TCACACACCTGCAGAGTCCGCTGAGCTCCTCTTCTGGAAGATCTCTG 1340
|||||
505 .....ValLysValLeuAspLysAspGlyLysGluLeuLysGlu...G 518
```



```

2192 AAGACAGAGAAAGAGAGAGAGGACGAAATGATTGAACCAAGAAAGAGC 2241
525 1ASnArGtLUeuVal 530
: 111 111
2242 AATAACAGAGCGCGTT 2257
seq_name: gb_gss::AZ674287

seq_documentation_block:
LOCUS      AZ674287      877 bp      DNA      GSS      14-DEC-2000
DEFINITION EMTK933TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION  AZ674287
VERSION    AZ674287.1  GI:11811433
KEYWORDS   GSS.
SOURCE      Entamoeba histolytica.
            Entamoeba histolytica.
ORGANISM   Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1 (bases 1 to 877)
AUTHORS     Loftus,B., Van Aken,S. and Fraser,C.
TITLE       Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library
COMMENT     Unpublished (2000)
            Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: entae@ligr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Reverse
            Class: Shotgun
            High quality sequence start: 18
            High quality sequence stop: 819.
            Location/Qualifiers
              1..877
                /organism="Entamoeba histolytica"
                /strain="HM1:IMSS"
                /db_xref="taxon:5759"
                /clone_id="Entamoeba histolytica Sheared DNA"
                /note="Vector: pHOSt1; site_1: Bst I; Constructed at The
                Institute for Genomic Research (TIGR), Rockville, MD.
                Genomic DNA isolated from broth cultures of E. histolytica
                using a method described by Clark and Diamond (Clark,
                C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
                method for isolate identification. Exp. Parasitol.
                77:450.). The DNA was mechanically sheared to give a
                tight size distribution (~2 kb). The v + i method used for
                the library construction is described in detail in Smith,
                H.O. and Venter, J.C. (Making small insert libraries for
                whole genome shotgun sequencing projects. In Genome
                Sequencing: A Practical Approach, eds. M. Vaudin and B.
                Barrell, Oxford University Press, 1999)."
```

seq_documentation_block: 943 bp DNA GSS 06-JUL-2001
LOCUS CNS070GA
DEFINITION T3 end of clone AZ0A009C10 of library AZ0A from strain CBS 712 of

Kluveromyces marxianus, genomic survey sequence.
 AL423584 GI:12206778
 VERSION AL423584.1
 KEYWORDS Kluveromyces marxianus.
 SOURCE Kluveromyces marxianus.
 ORGANISM Kluveromyces marxianus.
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Kluveromyces.
 REFERENCE 1 (bases 1 to 943)
 AUTHORS Llorente,B., Malpertuy,A., Blandin,G., Artiguenave,F., Wincker,P.
 and Dujon,B.
 TITLE Genomic exploration of the hemiascomycetous yeasts: 12.
 Kluveromyces marxianus var. marxianus
 FEBS Lett. 487 (1), 71-75 (2000)
 JOURNAL MEDLINE 20584712
 REFERENCE 2 (bases 1 to 943)
 AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
 Bolotin-Fukuhara,M., Bon,E., Brotlier,P., Casaregola,S.,
 de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
 Malpertuy,A., Neugeilise,C., Oziér-Kalogeropoulos,O., Potier,S.,
 Saurin,M., Tekala,F., Toffano-Mloche,C., Wesolowski-Louvel,M.,
 Wincker,P. and Weissenbach,J.
 TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of
 yeast species for molecular evolution studies
 FEBS Lett. 487 (1), 3-12 (2000)
 JOURNAL MEDLINE 20584711
 REFERENCE 3 (bases 1 to 943)
 AUTHORS Direct Submission
 TITLE Submitted (08-SEP-2000) Genoscope - Centre National de Séquençage,
 2 rue Gaston Cremieux, CP 5706, 91057 Evry cedex, FRANCE. (E-mail :
 seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 COMMENT This GSS is part of a random genomic sequencing program of thirteen
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
 exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,
 Saccharomyces kluyveri, Kluveromyces thermotolerans, Kluveromyces
 lactis var. lactis, Kluveromyces marxianus var. marxianus, Pichia
 angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
 5 kb were prepared and both extremities were sequenced. See
 keywords for description of this sequence and for the sequence of
 the other extremity of this insert.
 FEATURES
 source Location/Qualifiers
 1..943
 /organism="Kluveromyces marxianus"
 /strain="CBS 712"
 /varietal="marxianus"
 /db_xref="taxon:4911"
 /clone="AZ0AA009C10"
 /clone_11b="AZ0AA"
 /note="end : T3"
 <1..>933
 /note="similar to Saccharomyces cerevisiae ORF YJL176c [
 SW13 : transcription regulatory protein]
 1 putative frameshift(s)"
 /evidence="not_experimental"
 BASE COUNT 388 a 188 c 208 g 159 t
 ORIGIN
 alignment_scores:
 Quality: 131.50 Length: 414
 Ratio: 0.678 Gaps: 18
 Percent Similarity: 46.860 Percent Identity: 22.464
 alignment_block:
 US-09-754-947-1 x CNS070GA ..
 Align seg 1/1 to: CNS070GA from: 1 to: 943
 128 LysAspLeuGluThrLeuAsnTrpGly.....LysG1 138
 |||||||
 1 AAGACCTTGCGGTAAATTGGTTCCCATATGCCATTCAATAATATC 50

138 ULysAlaAsnIleLeuValGluLeuGlyIleSerValGlyThrGlyAsp. 154
 : |||||
 51 TGATATCAACGATTATGAGCACTTGGCATCTTGATGCGTCTGGTAAATC 100
 155GlnTrpGluProLysThrValThrLysAlaGluAla 167
 |||
 101 CTAAATATTTGCCAAGAAATTGCATCAAGAGCGCAATCATGATTCAGAGAG 150
 168 AlagInPheIleAlaLysThrAspLysGln.....PheGlyThrG1 181
 ||| |||||||
 151 AAGCAA.....AAGACTGATACGAGGTAAACCATATTCTTACGGA 191
 181 uAlaAlaLysValGluSerAlaLysAlaValThrThrGlnLysValGluV 198
 |||
 192 AGAAGAAAGAAATACAGAACTAAACAACGTGCGAGAAAGACGACG 241
 198 alLysPheSerLysAlaValGluLysLeuThrLysGlu.....AspIle 212
 |||
 242 ACAAGGAGGAGGACGAGAACAGAACAGACGAGGAGAAACAGACGAG 291
 213 LysValThrAsnLysAlaAsnAspLysValLeuValLysGluValThr 229
 ||| |||||
 292 AAGGAGCAGAAAGGCGGAGAGAAAGAAAGAAAGAAAGATATGCA 341
 229 rLeuSerGluAspLysArgSerAlaThrValGluLeuThrSerAsnLeuA 246
 :||| ||||| :||| ||| |||
 342 AATG...GTTGATTAACCTGAAGAAAGCAAGAGAG..... 372
 246 lAlaLysGlnThrThrThrValAspValAsnLysValGlyLysThrGlu 262
 :|||
 373 ..GCCGAAATGTAAACGATTCCAGACGTCAGAAAGCGATGATCTAAA 420
 263 ValAlaValGlySerLeuGluAlaLys...ThrIleGluMetAlaAspG1 278
 ||| |||||||
 421 GAACAAGAAATTCATTCATGAGCAAAAGAAAGAGCTGAAATGCTGATCA 470
 278 nThrValValAlaAspGluProThrAlaLeuGlnPheThrValLysAspG 295
 | |||||
 471 A.....GACCAAAAGCC...CAGACTCTACATCAAGATA 502
 295 LuAsnGlyThrGluValValSerProGluLysIleGluPheValThrPro 311
 :|||
 503 GAAT.....AGCCCAAGAGATAGAA..... 525
 312 AlaAlaGluLysIleAsnAlaLysGlyGluIleThrLeuAlaLysGlyTh 328
 |||||
 526AAATCCAC 533
 328 rSerThrThrValLysAlaValThrLysLysAspGlyLysValAlaLag 345
 |||||
 534 ATCTCCAAAGGCCAATGACGTCGAGAAAGAAAT..... 567
 345 lSerLysGluValLysValSerAlaGluLysAlaValAlaLaserIle 361

 567 567
 362 SerAsnTrpThrValAlaGluGlnAsnLysAlaAspPheThrSerLysAs 378

 567 567
 378 PheLysGlnAsnAsnLysValThrGluGlyAspAsnAlaThrValGluV 395
 |||||
 568AATATGAAAGCCAAAGATGAAAGCCAAAGCTGATTTGAAA 607
 395 alGluLeuLysAspGlnPheAsnAlaValThrThrGlyLysValGluTrp 411
 :|||
 608 CTCAGACTAAGACTGAACCGAAACTGAACCTCAACCGAAACCCAAAC 657
 412 GluSerLeuAsnThrGluValAlaValAspLysAlaThrGlyLysVal 428
 ||| |||||
 658 GAA.....ACCGAATCAAGCCCGTATCTGAAGCT..... 687

[illegible]

```

393 aGlnValGluLeuLysAspClnPheAsnAlaValThrThrGlyLysVal 409
      ::::::::::: ::::::::::: ||| :::::::::::
918 CGAGCGCTGAAGTAACAGAGAA.....GTTGAAGATGCCAAGAG 958
      ||| ::::::::::: :::::::::::
410 GluTyrGluSerLeuAsnThrGluValAlaValAlaLysAlaThrGln 426
      ||| ||| ::::::::::: :::::::::::
959 GAAGATGAGAGAGAGAAACAGAA.....AA 984
      :::::::::::
426 yLysValThrValLeuSerAlaGlyLysAlaProValLysValThrVal 443
      ::::::::::: :::::::::::
985 TTCTCTGCCAATACCTGCAGAGTAAGCGTGCAATCTCTGCACCTGATATCG 1034
      :::::::::::
443 yAspSerLysGlyLysAlaLeuValSerHisThrValGluIleGluAla 459
      :::::::::::
1035 CCAGCTCGAAGAGAGTAACCTTTCACAAATGCTGTATTGAGAGTCT 1084
      :::::::::::
460 PheAlaGlnLysAlaMetLysAspIleLysLeuGluLysThrAsn...Va 475
      :::::::::::
1085 CTCACAGAAACAGAAAGTAAGTAAGTCTGCAAGTAATCACTCAACAGCAA 1134
      :::::::::::
475 lAlaLeuSerThrLysAspValThrAspLeuLysValLysAla..... 489
      :::::::::::
1135 AATCTTAATCAAAAAACCCACCTGATGAACCTGAAAAAGCTGTGAGAG 1184
      :::::::::::
489 ..... 489
1185 ATATTAAATGAACATATTACCGATGCTCAGTTAGAAGCAATGACTGAACCTC 1234
      :::::::::::
490 .....Prova 491
1235 CATGACAGAACAGCAGTAATCAAGAGAAATGAAGAGAGAGAGGCCAA 1284
      :::::::::::
491 lLeuAspGlnTyr..... 495
      :::::::::::
1285 GCTTGAAATCTGCCTGCACAGAAAGACCAAGAACTGTGACATTATT 1334
      :::::::::::
496 .....GlyLysGluPheThrAlaProValThrValLysVal... 507
      :::::::::::
1335 CAGTCAGTGAAGAGAAAGAGATATATATATATTAACCTTAGAAAACAAAAT 1384
      :::::::::::
508 .....LeuAspLysAspGlyLysGluLeuLysGluGln 518
      :::::::::::
1385 ATTGACATATCTAAATCTGAGAGAGAAAGATCAGAGAAAGCAACA 1434
      :::::::::::
518 nLysLeuGluAlaLysTyrValAsnArgGluLeuValLeuAsnAlaAlaG 535
      :::::::::::
1435 GATGTTTGAAATTAAGCTTATTAATCTGAGAAATTAAGATACTATT 1484
      :::::::::::
535 LysGlnGluAlaGlyAsnTyrThrValValLeuThrAlaLysSerGlyGlu 551
      :::::::::::
1485 TGCAAACAGTAGAT.....TTAGTTCTCAAGAGACTGTGAGAA 1522
      :::::::::::
552 LysGluAlaLysAlaThrLeuAlaLeuGluLeuLysAlaProGlyAlaPhe 568
      :::::::::::
1523 AAAGAGCA..... 1531
      :::::::::::
568 eSerLysPheGluValArgGlyLeuAspThrGlu.....LeuAspLys 583
      :::::::::::
1532 .....AATATTCAAGCAGTTGATAGTAGAGTTGGCTTACAAGG 1571
      :::::::::::
583 yrValThrGluGlu.....AsnGlnLysAsnAla 592
      :::::::::::
1572 AAGACACCCCAAGAAATTTGGGAGAGACGACAAAAAGATGCTG 1621
      :::::::::::
593 MetThrValSerValLeuProValAspAlaAsnGlyLeuValLeuLysGln 609
      :::::::::::
1622 .....ATCAGCAATACAAAGTGTGATAGCAAC 1650
      :::::::::::
609 yAlaGluAlaAlaGluLeuLysValThrThrThrAsnLysGluLysGln 626
      :::::::::::
1651 ATGTGAGGACACAGATGTG.....GCTCAGA 1676
      :::::::::::
626 lValAlaSerAlaThrAspAlaGlnValThrValGlnAsnAsn.....Ser 640

```

```

1677 AAGTGATGAAGACAGTCTGATGATACGACAGACTAATGATGGAGAA 1726
      ::::::::::: ::::::::::: ||| :::::::::::
641 ValIleThrValGlyGln.....GlyAlaLys 649
      :::::::::::
1727 GTGCTCCAGATAGGCCAGAAATTAATTAATTAAGCCCATGGTGGTCTCTGA 1776
      :::::::::::
649 sAlaGlyGluThrTyrLysValThrValValLeuAspGlyLysLeuIle 666
      :::::::::::
1777 GCGTGGTGGTACTAAGAGAGTTCTATTAA..... 1807
      :::::::::::
666 hThrHisSerPheLysValValAspThrAlaProThrAlaLysGlyLeu 682
      :::::::::::
1808 .....GAATAGTGAATGAATGAATGAATGAAGAGT... 1840
      :::::::::::
683 AlaValGluPheThrSerThrSerLeuLysGluValAlaProAsnAlaAs 699
      :::::::::::
1841 .....AAAAATAGGACACACCAATAAACACTTC 1869
      :::::::::::
699 pLeuLysAlaAlaLeuLeuAsnIleLeuSerValAspGlyValProAla 716
      :::::::::::
1870 A.....GAGAACATATGACATCATGAGAGAACACAGAA 1904
      :::::::::::
716 hThr.....AlaLysAlaThrAlaSerAsn 724
      :::::::::::
1905 CAACTGAAGGTGAAGAAATCACTGATCAAGTACAGTGAAGAAATGAG 1954
      :::::::::::
725 ValGluPheValSerAlaAspThrAsnValValAlaGlnAsnGlyThrVa 741
      :::::::::::
1955 GTCGAAGGTGGTGGCTGATACTGACCAAAAGCT.....TT 1992
      :::::::::::
741 lGlyAlaLysGlyAlaThrSerIleTyrValLysAsnLeuThrValVal 758
      :::::::::::
1993 AGGAAGTGAA.....GTT 2006
      :::::::::::
758 yAspGlyLysGluGlnLysValGluPheAspLys 769
      :::::::::::
2007 AGGATGCTTCTAAAGTCACCTACTCAGATAGATAA 2041
      :::::::::::

```


OM of: US-09-754-947-1 to: N_Geneseq_1101.* out_format : pfs
Date: Mar 23, 2002 9:17 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+g2n.model -DEV=xlp
-O=/cgn2_1/USPTO.spool/US09754947/umat_22032002.125445-25485/app-query.fasta.1.854
-DB=N_Geneseq_1101 -OFMT=fastap -SUFFIX=ring -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -YCAPOP=10.000 -YGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YCAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsum62
-TRANS=trans40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09754947@cgn2_1.361 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPRX
-WAIT -THREADS=1

Search information block:

Query: US-09-754-947-1
Query length: 785
Database: N_Geneseq_1101.*
Database sequences: 930621
Database length: 428662619
Search time (sec): 173.620000

Score list:

Sequence	Strid Orig	ZScore	EScore	Len	Documentation	
/SID2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH28330			3875.00	4873.00	2.4e-263	2370
/SID2/gcgdata/geneseq/geneseqn/NA1997.DAT:AAH75488			419.50	517.17	9.9e-21	2766
/SID2/gcgdata/geneseq/geneseqn/NA1997.DAT:AAH71798			419.50	517.17	9.9e-21	2766
/SID2/gcgdata/geneseq/geneseqn/NA1999.DAT:AAH22751			415.00	511.51	2.0e-20	2763
/SID2/gcgdata/geneseq/geneseqn/NA1995.DAT:AAH99440			373.00	454.98	1.9e-17	4197
/SID2/gcgdata/geneseq/geneseqn/NA1997.DAT:AAH75487			248.50	299.21	1.4e-08	3687
/SID2/gcgdata/geneseq/geneseqn/NA1999.DAT:AAH71797			248.50	299.21	1.4e-08	3687
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAH71797			248.50	299.02	1.4e-08	3687
/SID2/gcgdata/geneseq/geneseqn/NA1999.DAT:AAH78246			248.50	299.02	1.4e-08	3687
/SID2/gcgdata/geneseq/geneseqn/NA1999.DAT:AAH78245			248.50	299.88	1.4e-08	3832
/SID2/gcgdata/geneseq/geneseqn/NA1999.DAT:AAH78245			248.50	299.61	1.9e-08	4088
/SID2/gcgdata/geneseq/geneseqn/NA1997.DAT:AAH75487			241.50	285.19	8.3e-08	6744
/SID2/gcgdata/geneseq/geneseqn/NA1996.DAT:AAH71476			240.50	283.25	1.1e-07	7291
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAH92489			238.50	280.78	1.5e-07	7253
/SID2/gcgdata/geneseq/geneseqn/NA1991.DAT:AAH11789			231.50	271.45	4.8e-07	7152
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAH65950			225.50	216.77	0.0005	183012
/SID2/gcgdata/geneseq/geneseqn/NA1996.DAT:AAH52184			224.00	269.32	6.3e-07	3388
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAH52184			224.00	269.27	6.4e-07	3306
/SID2/gcgdata/geneseq/geneseqn/NA1997.DAT:AAH90993			221.50	262.91	1.4e-06	4803
/SID2/gcgdata/geneseq/geneseqn/NA1993.DAT:AAH95511			218.50	259.31	2.3e-06	4702
/SID2/gcgdata/geneseq/geneseqn/NA1994.DAT:AAH72286			214.50	250.91	6.7e-06	6342
/SID2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH59105			214.00	250.43	4.3e-06	4287
/SID2/gcgdata/geneseq/geneseqn/NA1994.DAT:AAH72285			213.00	252.41	5.7e-06	4784
/SID2/gcgdata/geneseq/geneseqn/NA1997.DAT:AAH72285			212.50	250.81	4.1e-06	3294
/SID2/gcgdata/geneseq/geneseqn/NA1992.DAT:AAH24123			211.50	250.65	6.9e-06	4612
/SID2/gcgdata/geneseq/geneseqn/NA1997.DAT:AAH74374			211.50	245.75	1.3e-05	8155
/SID2/gcgdata/geneseq/geneseqn/NA1999.DAT:AAH75933			209.00	250.11	7.4e-06	3945
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAH52178			209.00	251.42	6.2e-06	2910
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAH52178			208.00	249.22	8.3e-06	3263
/SID2/gcgdata/geneseq/geneseqn/NA1993.DAT:AAH95507			208.00	245.66	1.3e-05	4937
/SID2/gcgdata/geneseq/geneseqn/NA1997.DAT:AAH90995			208.00	245.66	1.3e-05	4937
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAH52197			208.00	245.65	1.3e-05	4940
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAH52196			207.00	244.50	1.0e-05	3443
/SID2/gcgdata/geneseq/geneseqn/NA1993.DAT:AAH95507			207.00	244.09	1.6e-05	5116
/SID2/gcgdata/geneseq/geneseqn/NA1994.DAT:AAH72293			207.00	244.09	1.6e-05	5116
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAH52195			206.00	242.83	1.9e-05	5116

/SID2/gcgdata/geneseq/geneseqn/NA1997.DAT:AAH90996	206.00	237.80	3.6e-05	9171
/SID2/gcgdata/geneseq/geneseqn/NA1993.DAT:AAH95508	206.00	237.76	3.6e-05	9220
/SID2/gcgdata/geneseq/geneseqn/NA1996.DAT:AAH36122	205.00	242.00	2.1e-05	4865
/SID2/gcgdata/geneseq/geneseqn/NA1999.DAT:AAH20196	204.00	244.10	1.6e-05	3294
/SID2/gcgdata/geneseq/geneseqn/NA1999.DAT:AAH20178	204.00	244.10	1.6e-05	3294

seq_name: /SID2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH28330

seq_documentation_block:

ID AAH28330 standard; DNA; 2370 BP.

AC AAH28330;

DT 05-SEP-2001 (first entry)

DE Nucleotide sequence of a bacterial surface array protein (SAP).

KW Surface array protein; SAP; bacterial detection; ss.

OS Bacillus anthracis.

XX Key Location/Qualifiers

FT CDS 1..2349

FT /product="surface array protein"

FT /note="the nucleotides encoding residues 1-2 of

FT the protein are not given"

PN WO200149823-A2.

PD 12-JUL-2001.

PF 04-JAN-2001; 2001WO-US00358.

PR 06-JAN-2000; 2000US-0174901.

XX (BIOS-) BIOSITE DIAGNOSTICS INC.

PI Lee BA, Flores BM, Valkits GE;

WP: 2001-418358/44.

P-PSDB; ABA84651.

PT Novel methods and kits for detecting the presence of Bacillus anthracis

PT In a test sample -

XX Disclosure; Page 61-62; 62pp; English.

CC The present sequence encodes a bacterial surface array protein (SAP).

CC SAP is used in the method of the invention. The specification describes

CC a method for detecting the presence of Bacillus anthracis in a test

CC sample. The method comprises contacting the sample with a capture

CC reagent and detecting whether the a SAP is bound to the capture

CC reagent, which is indicative of the presence of B. anthracis in the

CC sample. The method is useful for detecting the presence or absence of

CC B. anthracis in a test sample.

XX Sequence 2370 BP; 940 A; 380 C; 459 G; 591 T; 0 other;

XX alignment_scores: Quality: 3875.00 Length: 783

XX Ratio: 4.949 Gaps: 0

XX Percent Similarity: 100.000 Percent Identity: 100.000

XX alignment_block: US-09-754-947-1 x AAH28330 ..

XX Align seg 1/1 to: AAH28330 from: 1 to: 2370

3 lysrhphpeproaspvalproalaasphistpolsyllleaspsseileas 19

|||||CCCCCGACGCGTTCCTGCTGATCCTGCGGGAATTGATTCATTAA 50

19 nTyrLeuValGluLysGlyAlaValLysGlyAsnAspLysGlyMetPheG 36
 |||||
 51 CTACTTAGTAAAGAAAGCGCAGTTAAAGTAAAGCAAGCAAGATTTGG 100
 |||||
 36 LurProGlyLysGluLeuThrArgAlaGluAlaAlaThrMetCalaGln 52
 |||||
 101 ACCCTGGAAAAAGATTAACTCGTGCAAGCAGCTAACATGATGGCTCAA 150
 |||||
 53 IleLeuAsnLeuProIleAspLysAspAlaLysProSerPheAlaAsp 69
 |||||
 151 ATCTTAACCTTACCAATCGATTAAGATGCTAAACCACTTCCTGCTGACTC 200
 |||||
 69 rGlnGlyIntPyrThrProPheIleAlaAlaValGluLysAlaGlyV 86
 |||||
 201 TCAAGGCCAATGTGTAACATCCATTCATCGCAGCTGAGAAAAAGCGCG 250
 |||||
 86 aLIleLysGlyThrGlyAsnGlyPheGluProAsnGlyLysIleAspArg 102
 |||||
 251 TTATTAAAGGTACAGGAAACGGCTTTGACCAACGGAAAAATCGACCCG 300
 |||||
 103 ValSerMetAlaSerLeuLeuValGluAlaIleGlyLysLeuAspThrLysVa 119
 |||||
 301 GTTCTATGCGATCTCTCTTCTTACAGCTTACAAATTAGATCTAAAGT 350
 |||||
 119 LAsnGlyThrProAlaThrLysPheLysAspLeuGluThrLeuAsnTrpG 136
 |||||
 351 AAACGGTACTCCAGCAACATAAATTCAGAGATTAGAAACATTAACTGGG 400
 |||||
 136 LysGluLysAlaAsnIleLeuValGluLeuGlyIleSerValGlyThr 152
 |||||
 401 GTAAAGAAAGAAAGCTAACATCTTAGTTGAATTAGAAATCTCTGGTACT 450
 |||||
 153 GlyAspGlnTrpGluProLysLysThrValThrLysAlaGluAlaAlaG 169
 |||||
 451 GGTGATCATGGGAGCCTAAAGAAACTGTAACAAAGCAAGAGCTGCTCA 500
 |||||
 169 nPheIleAlaLysThrAspLysGlnPheGlyThrGluAlaAlaLysValG 186
 |||||
 501 ATTCAATGCTTAAGACTGACAAACAGCTGCTGACAGAGCAGCAAAAGTTG 550
 |||||
 186 LysSerAlaLysAlaValThrThrGlnLysValGluValLysPheSerLys 202
 |||||
 551 AATCTGCAAAAGCTGTACACTCAAAAAGTAGAAGTTAAATTCAGCAAA 600
 |||||
 203 AlaValGluLysLeuThrLysGluAspIleLysValThrAsnLysAlaAs 219
 |||||
 601 GCTGTTGAAAAATTAACTAAAGAACATATCAAGTAACATAACAAGCTTA 650
 |||||
 219 nAsnAspLysValLeuValLysGluValThrLeuSerGluAspLysArgS 236
 |||||
 651 CAACGATTAAGTACTAGTTAAAGAGCTTAACCTTATCAGAACGATTAAGAAT 700
 |||||
 236 rAlaIleThrValGluLeuTyrSerAsnLeuAlaAlaLysGluThrThr 252
 |||||
 701 CTGCTACAGCTTGAATTATATGTAATTAAGCAGCTAAACAAACTTACACT 750
 |||||
 253 ValAspValLAsnLysValGlyLysThrGluValAlaValAlaGlySerLeuG 269
 |||||
 751 GTAGATGTAAACAAAGCTGTAAACACAGAGTGTAGCTTCTTTAGA 800
 |||||
 269 uAlaLysThrIleGluMetAlaAspGlnThrValAlaAlaAspGluProT 286
 |||||
 801 AGCAAAACCAATCGAAATGCGTACCAAAACAGTTGTAGCTGATGAGCCAA 850
 |||||
 286 hAlaLeuGlnPheThrValLysAspGluAsnGlyThrGluValValSer 302
 |||||
 851 CAGCAATTACCACTTCAAGTAAAGTGAAGAGGTAAGTGAAGTGTCTTCA 900
 |||||
 303 ProGluGlyIleGluPheValThrProAlaAlaGluLysIleAsnAlaLys 319
 |||||
 901 CCAGAGGCTATTGAATTTGTAACGCGCAGCTCGAGAAAAAATTAAATGCAAA 950

319 sGlyGluIleThrLeuAlaLysGlyThrSerThrThrValLysAlaValT 336
 |||||
 951 AGCTGAATACACTTTAGCAAAAGGTACTCACTCACTACTGTAATAAGCTGTT 1000
 |||||
 336 YrLysLysAspGlyLysValValAlaGluSerLysGluValLysValSer 352
 |||||
 1001 ATAAAAAACACGCTAAAGTACAGTGAAGGTGAAGAAAGTAAGTAAGTTCT 1050
 |||||
 353 AlaGluGlyAlaAlaValAlaSerIleSerAsnTrpThrValAlaGluG 369
 |||||
 1051 GCTGAAGGTGCTGCAAGTACCTCAATCTTACATGAGCAGCTGACAGAAC 1100
 |||||
 369 nAsnLysAlaAspPheThrSerLysAspPheLysGlnAsnAsnLysValT 386
 |||||
 1101 AAATTAAGCTGACTTACTTAAAGATTTCAAACAAAAACAATAAAGTTT 1150
 |||||
 386 YrGluGlyAspAsnAlaIleValGlnValGluLeuLysAspGlnPheAsn 402
 |||||
 1151 ACGAAGGCGCACACGCTTACGTTCAAGTAGAATGAAAGATCAATTTTAC 1200
 |||||
 403 AlaValThrThrGlyLysValGluThrGlyIleSerLeuAsnThrGluValAl 419
 |||||
 1201 GCAGTAACACTGGAAGAAAGTTCAATATGAGTCGTTAAACACAGAACTTGC 1250
 |||||
 419 aValValAspLysAlaThrGlyLysValThrValLeuSerAlaGlyLysA 436
 |||||
 1251 TGTAGTAGTAAGCTACTGGTAAGTAACTGATATCTGACAGGAAG 1300
 |||||
 436 LProValLysValThrValLysAspSerLysGlyLysAlaLeuValSer 452
 |||||
 1301 CACCGTAAGAACTAACTGTAAGAAATTCAAAAGTAAGTAAGCACTGTTTCA 1350
 |||||
 453 HisThrValGluIleGluAlaPheAlaGlnLysAlaMetLysAspIleLys 469
 |||||
 1351 CACACAGTTGAATGTGAAGCTTCCGCTCAAAAAGCAATGAAGAACATTTAA 1400
 |||||
 469 sLeuGluLysThrAsnValAlaLeuSerThrLysAspValThrAspLeuL 486
 |||||
 1401 ATTAGAAAAAACTAAACGTAGCGCTTTCTTCAAAAAGATGTAAACAGATTAA 1450
 |||||
 486 YsValLysAlaProValLeuAspGlnTyrGlyLysGluPheThrAlaPro 502
 |||||
 1451 AAGTAAGAGCTCCAGTACTAGATCAATACGGTAAGAGCTTTACACTCTCT 1500
 |||||
 503 ValThrValLysValLeuAspLysAspGlyLysGluLeuLysGluGlnLys 519
 |||||
 1501 GTAACAGTGAAGTACTGTGATTAAGATGTGAAGATTTAAAGAACAA 1550
 |||||
 519 sLeuGluAlaLysTyrValAsnArgGluLeuValLeuAsnAlaAlaGlyG 536
 |||||
 1551 ATTAGAACCTAAATATGTGAACACAGAAATTGATTCTGAATGCAGCAGGTC 1600
 |||||
 536 LngLuAlaGlyAsnPyrThrValValLeuThrAlaLysSerGlyGluLys 552
 |||||
 1601 AAGAGCTGCTGATTAATACAGTGTGATTAACCTGCAAAATCGGTGA 1650
 |||||
 553 GluAlaLysAlaThrLeuAlaLeuGluLeuLysAlaProGlyAlaPheSe 569
 |||||
 1651 GAAGCAAAAGCTACATTAGCTCTGAATTAAGAGCTCCAGGTGCACTTCTC 1700
 |||||
 569 rLysPheGluValArgGlyLeuAspThrGluLeuAspLysTyrValThrG 586
 |||||
 1701 TAAATTTGAAGTTCGTGCTTGAACACAGAAATTGATTAATATGTACTG 1750
 |||||
 586 LngLuAsnGlnLysAsnAlaMetThrValIleThrValLeuProValAspAla 602
 |||||
 1751 AGGAAACCAAAAGAAATGCAATGACTGTTTCAAGTTCTTCCGTGATGACA 1800
 |||||
 603 AsnGlyLeuValLeuLysGlyAlaGluAlaAlaGluLeuLysValThrTh 619
 |||||
 1801 AATGATATTAGATTAAAGGTGCAAGAACACAGCTGACATAAAGATACAC 1850
 |||||
 619 rThrAsnLysGluGlyLysGluValAspAlaThrAspAlaGlnValThrV 636

145 GtLueGly11LeSerValGltYrthGlyaspGlnTrpGluProValysrTh 161
::: ||||| ||| |||:::
526 CACCAACCAAGCTTCGGTCGATACC.....AAAA 554
161 rVaLTrhLySaLaGluAlaLaGlnPheLLeAlaYsrThrAspLySgInP 178
::: |||||::: |||||::: |||||::: |||||:::
555 CACCACTCTCCGCTGACTTTTGGCCCAATTTGTAAATAGAGCGGTAATTT 603
178 heGlyrhcGuaAlaLaLysValaGluSerAlaLysAlaYsrThrGln 194
::: |||||::: |||||::: |||||::: |||||:::
604AATCAGTCCGCAAGAATAAGTTGAAGTAACGCCGTTAATTCGACT 648
195 LysValaGluValaLysPheSerLysAlaValaGluLysLeuThrLysGluAs 211
|||::: |||::: |||::: |||:::
649 ACAGTGAATCAATTCATCAATCCGCAATTT.....GCTGA 683
211 pIleYsrValTrhAsnLyAlaAsnAsnAspLysValaLeuValLysGluV 228
|||::: ||||| |||::: |||::: |||::: |||:::
684 TGTGTGATTTACCAAAATTTTGGTCATCGATAACGGTTAACTGTTACTAAAG 733
228 aLrThrLeuSerGluAspLysArgSerAlaTrhValaGluLeuTrpSerAsn 244
::: |||||::: |||||::: |||||::: |||||:::
734 CAACCTCTTCTCGTGTATATAAAATCCGTAAGCTGTGCTGAATTAACCG 783
245 LeuAlaLaLysGlnTrhTyTrhValaLysPyaLAsnLysValGly.... 259
784 TTTTACCTCGTAATCAGCAAAATATCAATTAACAGCGACGCAATTAATAATTT 833
260 ..LysThrGluValaLa.....ValGlySerLeu. 268
||| |||||::: |||||::: |||||::: |||||:::
834 AAAAGCGGACCCGCTTAAGCAATTAACGTGTAAGTTGTGTGCTGTTC 883
269 ..GluAlaLysThrLLeGluMetAlaAspGlnTrhValaValaLaAspGlu 284
::: |||||::: |||||::: |||||::: |||||:::
884 AAGATCCGTAACGTGTGCACTAAATTAATAGTTCCTTAAGGTTGGAAG 933
285 ProThrAlaLeuGlnPheThrValaLysAspGluAsnGlyTrhGluVala 301
|||::: |||::: |||||::: |||||::: |||||:::
934 GAATCGCTTTA.....ACTGTAAAGATCGAGTGGCAAAAGATGTTG 977
301 L.....SerProLueLyLLeGluPheValT 310
978 AGGTGCTAAAGTAGAATTAATCTTCTTCTTAATCTAATTAATTTGTTAGTTT 1027
310 hPProAlaLaGluLysLLeAsnAlaLysGlyGluLLeThrLeuAlaLys 326
::: |||||::: |||||::: |||||::: |||||:::
1028 CAAGTCCGCAATATCAGATATCTGCTGCTAAAGTAACAGCTGTAAACCG 1077
327 GlyTrhSerTrhValaLysAlaValTyryLysLysAspGlyLysVala 343
|||||::: |||||::: |||||::: |||||:::
1078 GGAAAGCGTGAATTAAGTTCGCAAAAGTTCAATTAACAGATGCTGTGTTACT 1127
343 lAlaGluSer.....LysGluValaLysSerAlaGlu 355
::: |||||::: |||||::: |||||::: |||||:::
1128 AACCAATACATTTAAAGTACAGATTACAGAAGTGGCTGTCAAGTCCAAA 1177
355 lValAlaValaLaserLLeSerAsnTrpTrhValaLaGluGln..... 369
::: |||||::: |||||::: |||||::: |||||:::
1178 ATCAAGGATTAATCTTAAGTATATCTTTCTAATGCTCCACAGATTAACA 1227
370AsnLysAlaAspPheThrSerLysAspPheLysGlnAsnAs 383
|||||::: |||||::: |||||::: |||||:::
1228 GTTGCAATTAACAAAGCTGAGAAAGTAACTTCAATGTTTCTGGAGAAAC 1277
383 nLys.....ValTyryGlu.....GlyAspAsnAlaTyryValG 394
|||::: |||||::: |||||::: |||||:::
1278 TAAACACAGTTCGAATGTATGATACTAAAAAGGATGATCTGAACATAAC 1327
394 lValaGluLeuLysAspGlnPheAsnAlaValTrhThrGlyLysValaGlu 410
|||::: |||||::: |||||::: |||||:::
1328 GTGTGATTTTCAAAAT.....GCACAT 1350

[illegible]


```

2155 ..GTAGCTGGA..... 2163
675 rAlapThrValLysGlyLeuAlaValGluPheThrSerThrSerLeuL 692
2164 .....ACAGCAGACAGAAATGACA.....A 2182
692 ySGluValAlaProAlaAspLeuLysAlaAlaLeuLeuAsnIleLeu 708
2183 AACGGATT...CCAGGGAAATATATCAACTGCA..... 2214
709 SerValAspGlyValProAlaThrThrAlaValAlaThrAlaSer...As 724
2215 .....AGTGTAACCTGCAGCGAACAAGTAGCAACAGCTGTTAC 2255
724 nValGluPheValSerLysAspThrAsnValAlaIleAsnGlyThrV 741
2256 TGTAAAGCTTAATTCAAGTCATATGATTTACATTGGAAGATTAAAT 2305
741 aLgylAlaLysGlyAlaThrSerIleThrValLysAsn..... 753
2306 TCGGTGTAATTCAGCCATCAACAATTA...GTCAACATCAACACATCAAC 2352
754 .....LeuThrValValLysAspGlyLysGluGlyValGluPheAs 768
2353 GAATTTATTCGACGTTTCAAAAGCGGCTAAATAAGATGATTTCTATA 2402
768 pLysAlaValGluValAlaValSerIleLysGluAla 780
2403 TAAACCGCTT.....GTAAAGGTTAAAGATGCA 2430

seq_name: /SID2/gcgdata/geneseq/geneseq/NA2000.DAT:AAA71798
seq_documentation block:
ID AAA71798 standard; DNA; 2766 BP.
XX
XX AAA71798:
XX AC
XX 08-JAN-2001 (first entry)
XX DE
XX B. stearothermophilus sbsb DNA.
XX
XX S-layer gene: sbsb; antibacterial; vaccine; adjutant; bioreactor;
XX KM polyhydroxyalkanoate synthesis; ds.
XX OS
XX Bacillus stearothermophilus.
XX
XX Key Location/Qualifiers
XX CDS 1..2766
XX FT /tag= a
XX FT /product= "sbsb"
XX FT sig_peptide 1..93
XX FT /tag= b
XX mat_peptide 94..2763
XX FT /tag= c
XX
XX DE19903345-A1.
XX
XX 03-AUG-2000.
XX
XX 28-JAN-1999; 99DE-1003345.
XX
XX 28-JAN-1999; 99DE-1003345.
XX
XX (LUBI/) LUBITZ W.
XX
XX Lubitz W;
XX PI
XX WPI: 2000-533868/49.
XX DR
XX P-PSDB; AAB10626.
XX
XX Host cell, useful e.g. as bioreactor for production of
XX poly(hydroxyalkanoate), containing two or more recombinant
XX

```

```

PT polypeptides, with at least one in carrier-bound form -
XX
XX Disclosure, Page 16-19; 26pp; German.
XX
CC This invention describes a novel host cell (A) comprising
CC functional recombinant polypeptides (1), at least one being
CC bound form. The products of the invention have antibacterial
CC (A), or, where bacterial, their ghosts (B), are useful as
CC adjuvants (specifically for presentation of immunogenic ep
CC pathogens or autologous immunostimulation polypeptides), e.
CC or preferably, as enzyme reactors for performing a cascade
CC specifically synthesis of poly(hydroxyalkanoate). Localiza
CC individual (1), specifically enzymes, in separate cellular
CC avoids adverse reactions between products and substrates,
CC as bioreactors. (1) can be produced in carrier-bound form
CC function. This sequence encodes the Bacillus stearotheroph
CC protein sbbs which is used to illustrate the method of the
XX
XX
SQ Sequence 2766 BP; 996 A; 465 C; 534 G; 771 T; 0 other:

alignment_scores:
      Quality: 419.50      Length: 879
      Ratio: 0.891      Gaps: 40
      Percent Similarity: 53.584      Percent Identity: 24.005

alignment_block:
US-09-754-947-1 x AAA71798 ..

Align seg 1/1 to: AAA71798 from: 1 to: 2766

4 ThrPheProAspValProAlaAspHisTrpGlyIleAspSerIleAsnTy 20
:::|::| |::|::| ::::: |::|::|::|::|::|
97 AGCTTCACAGATGTTGGCCGCATAT.....AAAGATGCATGATTT 140
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
20 rLeuValGluLysGlyAlaValLysGlyAsnAspLysGlyMetPheGlu 37
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
141 CTTAAGTATCAACTGCTGCACAAAAGGTAAACACAAACAAAATTGCGG 190
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
37 roGlyLysGluLeuThrArgAlaGluAlaIleAlaIleAlaIleAlaIle 53
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
191 TTTCAGATGAATATCTGCTGTAGATGGCGCGATTATTTCTGCACAGTA 240
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
54 LeuAsnLeuProIleAspLysAspAlaLysProSerPheAlaAspSerG 70
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
241 TTAAACTATGACGTTGACACAGCAAAACGAGGCTTCACAGATGTGCC 290
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
70 ngLyGlnTrpTyrThrProPheIleAlaAlaValGluLysAlaGlyVal 87
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
291 AAAAGAC...CGTCCAAATACGTCACGCGCTGTAGAGCTGGCGTAT 337
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
87 lelysGlyThiGlyAsnGly...PheGluProAsnGlyLysIleAspArg 102
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
338 TAAACGGTTAAAGCACCTGGCAAAATTGGTCATACGACCCATTAATCTGC 387
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
103 ValSerMetAlaSerLeuLeuValGluAlaTyrLysLeuAsp..... 116
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
388 GTTGAATATGCAAAAATCATCGCGACGAGCTTACAAATTTAAAGCTGACGA 477
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
117 .....ThrLysValAsnGlyThrProAlaThrLysPheL 138
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
128 ysaSplLeuGluThrLeuAsnTrpPheLysGluLysAlaAsnIleLeuVal 144
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
438 TGTAAACTTCATCTACCTGATGTAAAGATGACATGCGACCATACGTAA 487
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
488 AAGCGCTT.....TATAATAACGAGTAACCAAAAGGTTAAAA 535
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
145 GluLeuGlyLysSerValGlyThrGlyAspGlnTrpGluProLysLysTh 161
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
526 CACCAACAAGCTTGCGTCCATAC.....AAAAA 554
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
161 rValThrLysAlaGluAlaIleAlaGlnPheIleAlaLysThrAspLysGln 178
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

555 CATCACTCTGCGGACTTTCGCAATTGTATATAGCGCGTAATATT. 603
178 heGlyThrGluAlaValLeuSerAlaValAlaValThrThrGln 194
    ::::: ::::: ::::: :::::
604 ....AATGCAGTGCAGAAATAGTTGAAGTAACTCGCGTTAATTCGACT 648
195 LysValGluValLysPheSerLysAlaValGluLysLeuThrLysGln 211
    ::::: ::::: :::::
649 ACAGTGAAAGTAACTCAATACGCAAAATT.....CCTGA 683
211 PileLysValThrAsnLysAlaAsnAsnLysValLeuValLysGluV 228
    ::::: ::::: ::::: :::::
684 TGTGTATTCACAAATTTGCTATGATACGTTTAAGTCTTACTACTAAAG 733
228 aLThrLeuSerGluAspLysArgSerAlaThrValGluLeuTyrSerAsn 244
    ::::: ::::: ::::: :::::
734 CAACCTCTTCGCTGATAAAAATCCGTAGAGTGTGGTAATAATAACCG 783
245 LeuAlaValLysGlnThrTyrThrValAspValAsnLysValGly.... 259
    ::::: ::::: ::::: :::::
784 TTTACTCGTAATCAGAAATATACAAATTACAGCAGACGCAATTAATAATT 833
260 .LysThrGluValAla.....ValGlySerLeu. 268
    ::::: :::::
834 AAAGCGCAGACCGCTAAGAAATTAAGTGAATTTGTTGCTGCTTC 883
269 ..GluAlaLysThrIleGluMetAlaAspGlnThrValAlaAlaAspGlu 284
    ::::: ::::: ::::: :::::
884 AAGATCGCTACTGTGCCTAAATATAGTTCCCTTAAGTTGAGAG 933
285 ProThrAlaLeuGlnPheThrValLysAspGluAsnGlyThrGluValVal 301
934 GAATCGGTTA.....ACTGTAAAGATCAGATGCGCAAGATGTTGT 977
301 L.....SerProGluGlyIleGluPheValT 310
    ::::: :::::
978 AGCTGCTAAGTAGAACTTCTTCTATCTAATATATTGTTGAGTTT 1027
310 hPrcAlaAlaGluLysIleAsnAlaLysGlyGluIleThrLeuAlaLys 326
    ::::: :::::
1028 CAAGTCGCAGATATCAGTATCTGCTAAGTTAACAGCTGTAAACCG 1077
327 GlyThrSerThrThrValLysAlaValTyrLysLysAspGlyLysValVal 343
    ::::: :::::
1078 GGAACAGCTGATGTACTGCAAAAGTTACCAAGATGCTGTGTACT 1127
343 AlaGluSer.....LysGluValLysValSerAlaGluGln.... 355
    ::::: :::::
1128 ACAATATACATTTAAAGTACAGACTTACAGAAAGTCTGTCAAGTCCAAA 1177
355 LysAlaValAlaSerIleSerAsnTTPThrValAlaGluGln..... 369
    ::::: :::::
1178 ATCAAGGATTTCTTGTATTAATCTTCTTAATGCTCCACAGAAATACA 1227
370 .....AsnLysAlaAspPheThrSerLysAspPheLysGlnAsnAs 383
    ::::: :::::
1228 GTTGCAATTTAACAAAGCTGAGAAAGTCAATGTTGCTGAGAAAC 1277
383 nLys.....ValTyrGlu.....GlyAspAsnAlaTyrValG 394
    ::::: :::::
1278 TAAACACATGTGCAATGTATGATACTAAAAACGGTATCTCGAACTAAC 1327
394 InValGluLeuLysAspGlnPheAsnAlaValThrThrGlyLysValGlu 410
    ::::: :::::
1328 CTGTGATTTCAAGAT.....GCAACT 1350
411 TyrGluSerLeuAsnThrGluValAlaValAlaLysLysAlaThrGlyL 427
    ::::: :::::
1351 GTACCTTATTAAATCAATATTGCAACACCTGCTATTAAATGCTAGACGA 1400
427 sValThrValLeuSer.....AlaGlyLysAlaProValLysV 440
    ::::: :::::
1401 GCTCTTCTCACAGCTAATGCTGCGCAATCTGGAAGAGCTTCATTTGAG 1450
440 aLThrValLysAspSerLysGlyLysAlaLeuValSerHisThrValGlu 456
    ::::: :::::
1451 TAAACATTAAGATATAAT.....ACAAAAGAACATTTTACA 1485
457 IleGluAlaPheAlaGlnLysAlaMetLysAspIleLysLeuGluLysTh 473
    ::::: :::::
1486 GTTGATGTAAAAAAGACCCCTGATTTACAAAGATATAAAGTGAATGCAAC 1535
473 rAsnValAlaLeuSer.....ThrLysAspV 482
    ::::: :::::
1536 TTCTGTTAACTTTCCGATGAAGCTGTGGCGCGGCAAGTGAAGAG 1585
482 aLThrAspLeuLysValLysAlaProValLeuAspGlnTyrGlyLysGlu 498
    ::::: :::::
1586 TTAACCAAAAAACGATTTAAAGTAAGTGCAGTTGACCAATACGCTAAAGA 1635
499 Phe.....ThrAlaPr 502
    :::::
1636 ATTTAAATTTGTACAAAAGTTAAAGTTACTGTACACTAATATACGAAG 1685
502 oValThrValLysValLeuAspLysAsp.....G 512
    ::::: :::::
1686 ACTAGTTATTTAAATGTAAATAGCGATTAATCAATTGACTTTGATACCG 1735
512 LysGluLeuLysGluGln..... 518
    :::::
1736 GCAATATGTCACACTGCACCAATTTGTTGCTTGCACAAACAAACAAATT 1785
519 .....LysLeuGluAlaLysTyrValAsnArgGluLeuValLeuAs 532
    ::::: :::::
1786 GTCAATGTAAAGTAGAAGTTAAATATTTTCAAAAT..... 1821
532 nAlaAlaGlyGlnGluAlaGlyAsnTyrThrValValLeuThrAlaLys 549
    ::::: :::::
1822 ...GCTAGTGCACACACACCACTTCACTAATAACATTAAGTTAATG 1867
549 eArgLysGluLysGluAlaLysAlaThr...LeuAlaLeuGluLeuLysAla 564
    ::::: :::::
1868 TAGTAATATGTAAGCTGACGCTACACCAAGTAGATTTGATATGTGACA 1917
565 ProGlyAlaPheSerLysPheGluValArgGlyLeuAspThrGluLeuAs 581
    ::::: :::::
1918 CCT.....TCPTAAATGATGTAAATGCTCCAAACACT..... 1950
581 pLysTyrValThrGluGluAsnGlnLysAsnAlaMetThrValSerVal 598
    ::::: :::::
1951 .GCTTCTATGACAGATGTTGATTTTATTAATTTGCAAAAGTGTGAGATT 1999
598 euProValAspAlaAsnGlyLeuValLeuLysGlyAlaGluAlaGlu 614
    ::::: :::::
2000 ACACACTCGATTAATGAT.....AGACGTCA 2028
615 LeuLysValThrThrThr.....AsnLysGluGlyLys 625
    ::::: :::::
2029 AAAAAAGTTACTCCAACTGCAACTACACTTGTAGTACAAAAAANA 2078
625 sGluValAspAlaThrAspAlaGlnValThrValGlnAsnSerValI 642
    ::::: :::::
2079 AAAAGTTAATGGAATGATTACAA.....TTCAAGGGACCAAGCAAT 2122
642 LeThrValGlyGlnGlyAlaLysAlaGluThrTyrTyrLysValThrVal 658
    ::::: :::::
2123 TAACGCTATCAACTCTCTAGTACAGGAAC..... 2154
659 ValLeuAspGlyLysLeuIleThrThrHisSerPheLysValAlaAspTh 675
    ::::: :::::
2155 ..GTAGATGA..... 2163
675 rAlaProThrAlaLysGlyLeuAlaValGluPheThrSerThrSerLeu 692
    ::::: :::::
2164 .....ACACAGAGAAGAAATGACA.....A 2182

```



```
488 AACGCTTTATTAATACGAATACAAAGGTAA..... 522
145 GluLeuGlyIleSerValGlyThrGlyAspGlnTrpGluProIleSylth 161
523 .....ACACACACAGCTTGGGTGACATACCAAA 551
161 ValThrLysAlaGluAlaGlnIlePheIleAlaLysThrAspLysGlnP 178
552 CACACACCGCGGTGACTTGGCAATTTGTATATAGACGCGTCAATATT. 600
178 heGlyThrGluAlaAlaLysValGluSerAlaLysAlaValThrTrgIn 194
601 .....AATGACAGTGCAGAAATAGTTGAAGTAACCTCGGTAAATTCGACT 645
195 LysValGluValLysPheSerLysAlaValGluLysLeuThrLysGlns 211
646 ACAGTAAAGTACATTCATACGCAATT.....GCTGA 680
211 PileLysValThrAsnLysAlaAsnAspLysValLeuValLysGluV 228
681 TGTGATTTCACAAATTTGCTATCGATTAACGGTTAACTGTTACTAAG 730
228 aLThrLeuSerGluAspLysArgSerAlaThrValGluLeuTrpSerAsn 244
731 CAACCTTTCTCGTATAAAAAATCCGTAGAGGTGTGTAATTAACCG 780
245 LeuAlaAlaLysGlnThrTyThrValAspValAsnLysValGly.... 259
781 TTTACTCGTATACGGAATATATCAATTACAGCAGCATTAATAATT 830
260 LysThrGluValAla.....ValGlySerLeu. 268
831 AAAAGCGGAGACCGCTAAGCAATTAACTGTAAGTTGTTGGTCTGTC 880
269 ..GluAlaLysThrIleGluMetAlaAspGlnThrValAlaAlaAspGlu 284
881 AAGATCCGTAACCTGTTGCACTAATAATATGTTCCGTTAAAGTTGAGAG 930
285 ProThrAlaLeuGlnPheThrValLysAspGluAsnGlyThrGluValVa 301
931 GAATCTGCTTTA.....ACTGTAAGAATCAGAGATGCCAAGATGTGT 974
301 L.....SerProGluGlyIleGluPheValT 310
975 AGGTGCTAAAGTAGAACTTACTTCTTAATCTAATATATTGTTGTGTTT 1024
310 hProAlaLeuGluLysIleAsnAlaLysGlyGluIleThrLeuAlaLys 326
1025 CAAGTGGCCAGATACGATCTGCTGCTAAAGTTACAGCTGTAACCG 1074
327 GlyThrSerThrThrValLysAlaValTyLysLysAspGlyLysValVa 343
1075 GGAAACAGCTGATGTACTGCAAAATTTACATTACAGATGTGTGTACT 1124
343 AlaGluSer.....LysGluValLysValSerAlaGlu 355
1125 AACCAATTAATTAAGTAGACATTACAGAGAGCTGTGCAAGTACAA 1174
355 ValAlaAlaValAlaSerIleSerAsnTrpThrValAlaGluGln..... 369
1175 ATCAAGAGATTACTTATGATTAATCTTCTATATGCTCCACAGATACA 1224
370 .....AsnLysAlaAspPheThrSerLysAspPheLysGluAsnAs 383
1225 GTTGATTTAAACAAGCTGAGAAGTAACCTCAATGTTGCTGAGAGAAC 1274
383 nLys.....ValTyGlu.....GlyAspAsnAlaTyValG 394
1275 TAAAAACAGTTGCAATGTATGATACATAAAACGCGTATCCTGAACATAAC 1324
394 lValGluLeuLysAspGlnPheAsnAlaValThrGlyLysValGlu 410
1325 CTGTTCATTTCAAAGAT.....GCAACT 1347

411 TyrgIleuAsnThrGluValAlaValAlaAspLysAlaThrGlyLy 427
1348 GATCGTTCATTAAATCCAAATTAATGCACACGCTGATTAAATGATAGGA 1397
427 sValThrValLeuSer.....AlaGlyLysAlaProValLysV 440
1398 GCTCCTTCACACAGCTAATGCTGGCCAAATCTGCAAAAGCTTCATTGGA 1447
440 aLThrValLysAspSerLysGlyLysAlaLeuValSerIleThrValGlu 456
1448 TACATTTTAAAGATTAAT.....ACAAAAGAACATTTTCA 1482
457 lIleGluAlaPheAlaGlnLysAlaMetLysAspIleLysLeuGluLysTh 473
1483 GTTGATGTGAAAAAAGACCCCTGATTACAGATTTAAAGTGAATGCAAC 1532
473 rAsnValAlaLeuSer.....ThrLysAspV 482
1533 TTTCTGTTAACTTTCGATGAAGCTGTTGCCGCGGGAAGTTGAAGAG 1582
482 aLThrAspLeuLysValLysAlaProValLeuAspGlnTyGlyLysGlu 498
1583 TTACCCAAAAAAGATTAAGTAAGTGCAAGTGCACCAATACGATAAGAA 1632
499 Phe.....ThrAlaPr 502
1633 ATTAATTTTGTACAAAAGGTAAAGTTACTGTTACACTAATATACGAAGG 1682
502 oValThrValLysValLeuAspLysAsp.....G 512
1683 ACTAGTTATTAAAAATGTAAATAGCATATATCAATTGACTTTGATAGC 1732
512 lLysGluLeuLysGluGln..... 518
1733 GCAATATGTCACCACTGACCAATTGTTGTGCTGTCACAAACAAAGCAAAAT 1782
519 .....LysLeuGluAlaLysTyThrValAsnArgGluLeuValLeuAs 532
1783 GTCAATGTGAAGTAGAAGTTAATATATTCAAAAT..... 1818
532 nAlaAlaGlyGlnGluAlaGlyAsnTyThrValValLeuThrAlaLys 549
1819 ...GCTAGTGACACACACCACTTCACTAAACAAATTAAGTTGTTAATG 1864
549 eGlyGluLysGluAlaLysAlaThr...LeuAlaLeuGluLeuLysAla 564
1865 TAGTGAATGTAAACCTGACGCTACACACGATGATTAATTTGTACGA 1914
565 ProGlyAlaPheSerLysPheGluValArgGlyLeuAspThrGluLeuAs 581
1915 CCT.....TCGAAATATGATGTGAATGCTCCAAACACT..... 1947
581 plysTyThrValThrGluGluAsnGlnLysAsnAlaMetThrValSerValL 598
1948 ..GCTTCTACTGAGATGTTGATTATTATTAATTTGCAAAAGTGTGATTT 1996
598 euProValAspAlaAsnGlyLeuValLeuLysGlyAlaGluAlaLeuGlu 614
1997 ATACACTGCAATCTTAATGTAAACGCTCTTAATAAAAGTACTACCAACTGCA 2046
615 LeuLysValThrThrAsnLysGluGlyLysGluValAspAlaThrAs 631
2047 ACTACACTGTGTGATTAAT.....GATTAATGTTGAAGTAAATG 2087
631 palGlnValThrValGlnAsnAsnSerValIleThrValGlyGlnGly 648
2088 GAATGATTTTACATTCACAGGTAACGATGAATTAACGTA.....T 2128
648 lAlaLysAlaGlyGlnThrTyThrLysValThrValValLeuAspLysLeu 664
2129 TAACCTTCTTCTAGTACAGTAACGTTGATGTACAGCTGATGGA..... 2172
```



```

912 AGTATGTTTAACTCTTGAC.....GGCAACAACATCGTGCTTTAA 955
244 AsnLeuAlaIAlaValSerLthrValAspValAsnLysValGly 260
956 GGCTGGCTGCT.....GTAGTCCCACTAAAGGTGACT 990
260 sThrGluValAlaVal..GlySerLeuGluAlaYsThrIleGluMet 276
991 AGTATCTCTCGACTCAAGGTAACTTGCTCAAGAAAGTAAAGTTCAAG 1040
276 LaspGlnThrValValAlaLaspGluProThrAla.....LeuGln 289
1041 CTAAAGTAACTGTGCTGAAGGTCAATCTAAAGCTGTATCTCTGTACT 1090
290 PheThrValLysAspGluAsnGlyThrGluValValSer..... 302
1091 TTCACGTGACCAAGTAAACAATGATGCGCTTGACCAACATTAAACAG 1140
302 ..... 302
1141 TGAAGCTTTAACAACAAGAGGTATCGCAACATCTCTACACTGTT 1190
303 ..ProGluGlyIleGluPheValThrProAlaAla..... 313
1191 ATAAAGAAAGTACTGATGAAGTAACTGTTATGCAACTGTGATCGTCT 1240
314 .....GluLysIleAsnAlaLys 319
1241 AAATTCTCACTGTTATGTTATTCGGGTGTAGATACAAATCTTTCACT 1290
319 sGlyGluIleThr.....LeuAlaLysGlyThrSerThrThrV 332
1291 TGAAGAAAGTAACTCAAGGTCTTCAATGATGTCGCAAAACAACCTT 1340
332 sLysValAlaValTyrLys.....LysAspGlyLysVal..... 342
1341 ACAAAAGTCTTATTAACAACTTAACGTGAACAGAACCAACAA 1390
343 .....ValAlaGluSerLysGly 348
1391 ACATTTAATGTTGGTTTGTAGAAAACATGATGTTACTCTGATTAAC 1440
348 uValLysValSerAlaGlyAlaAlaValAlaSerIleSerAsnTyrT 365
1441 ACCAAATGCTACAGTTAATGGCTAAAGCATTAACATTAACCAATG 1490
365 hValAlaIleGluLysAsnLysAlaAspPheThrSerLysAspPhe 381
1491 CAGCTTTAGAC.....GCTGCTCAAAATTACACACAGATTCTAAAG 1534
382 AsnAsnLysValTyrGluLysAspAsnAlaTyrValGlnValGlu 398
1535 GCTACATTCACAGTTTCTGTACTAATGACGTGAAGCCAGTAGTA 1584
398 sAsp.....GlnPheAsnAla 403
1585 TGATCTACACACCACTAACATAGTACTTCAAAATAAAATATAGT 1634
404 .....ValThrThrGlyLysValGluTyrGluSerLeuAsnThr 416
1635 CTGCTTTACAAACTACTGCTTAAAGTAACTTTCGCTGCTCTTCAAG 1684
417 GluValAlaVal.....ValAspLysAlaThrGlyLysValThr 429
1685 GAGTATACATATGAGTAACTGCTGCTGATATGCT...GGAGAAGT 1731
429 rValLeuSerAlaGlyLysAla..ProValLysValThrValLys 444
1732 TGCAATTCGCGCTACTAAGCGTGGCAATACAAAGTTATTGTAAGA 1781
445 .....SerLysGlyLysAlaLeuValSerHisThrValGln 456

```

```

1782 AAGCTTTAACTTACTTAAATAATGAATCGTTAATGTTGATTCATGA 1831
456 uIleGlu..... 458
1832 GATTAAGATCGTGTAAATTTCACCACTTACAAATGCTAAATTCGT 1881
459 .....AlaPheAlaGlnLysAlaMetLysAspIleLys..L 470
1882 TGATCCAGATACTGCGATATCTTACAGCGCATTAACCAAAACAA 1931
470 eGluLysThrAsn..... 474
1932 CTGTAAACAAATGATTAAGGTGAAGCTACATTTGTTATCGTTCTG 1981
475 .....ValAlaLeuSerThrLysAspValThrAspLe 485
1982 ACAGTAAAGATTAATGCAACACCACTTGTGATGATTAATTAAC 2031
485 uLysValLysAlaProValLeuAspGluTyrGlyLysGluPheThr 502
2032 TGATCAAAACCAAGCGACCTGATGAAGGTGAACCAAAAGCAGT 2081
502 rovalThr...ValLysValLeuAspLysAspGlyLysGluLeuLys 517
2082 CAATCTCTTACTCCAGCACCATATCTTGATGGCTCAGCTCAAAAC 2131
518 GlnLysLeuGluAlaLysTyrValAsnArgLysLeuValLeuAsn 534
2132 TACAAA.....AAATCAGATCTTAATAAGCTGTAACTAAGTTGA 2172
534 aGlyGlnGluAlaGlyAsnTyrThrValValLeuThrAlaLysSer 551
2173 TGGTCTCAAACTGCAGATTTTGCAGCAGAAATTAACCAAAAGCG 2222
551 LuLysGluAlaLysAlaThrLeuAlaLeuGluLeuLysAlaPro 567
2223 AAAAAGTACTGCTTCTTAT.....AAGAAAGCACT 2257
568 PheSerLysPheGluValArgLysAspThrGluLeuAspLysTyr 584
2258 TATACAAATCTCAATATCTGCTTAATGATTAATAAGTAGTAACCA 2307
584 lThrGluLysGlnLysAsnAlaMetThrValSerValLeuProVal 601
2308 TATCTCACCAATCTGCTACACAGTAACTTAAGAACT...ACTT 2354
601 sPalaAsnGlyLeuValLeuLysGlyAlaGluAlaGlu..... 614
2355 CTTCACAGGAAGTGTATTTACACCTGCTTAAGAAATTTGAAGTT 2404
615 .....LeuLysValThrThr 620
2405 GTGATGTTAAACAACGCTGTTAAACTAATGTGTAACAGTATTC 2454
621 AsnLysGluGlyLysGluValAspAlaThrAspAlaGlnValThr 636
2455 TAAATACAGACGTAAAGACTATGCAATTACTGTAAGAACTACAG 2504
636 ..... 636
2505 CATTCACAGCTACAAATGAAGTTCCAAACTCTTACCTGCGTAC 2554
637 .....GlnAsnAspSerValIleThrVal 644
2555 CAATTCAATACAGCTGATTCGTGTTCAAAACGCAACTATTTGGTT 2604
644 lGlyGln.....GlyAlaLysAlaGlyGluThrTyrLys 655
2605 TGTAATAAACCAAGTAAATATGCTGCTATCAGCGCAAAACATATA 2654
655 ..... 655
2655 ACTTCGAGCTAATGTAATGAAGTATTTGGTGAAGCGGATGGAGCA 2704

```

```

656 .....ValThrValVa 659
2705 TTATTAACTCAATATGCACTGAAGCCAAAAGTAACAAATCTCATATAA 2754
659 lLeuAspGlyLysLeuIleThrHisSerPheLysValValAspHis 676
2755 TGTAGATGGTATACAGTACATA.....TTTAAAGTATTATGT.... 2791
676 lAprThrAlaLysGlyLeuAlaValGluPheThrSerThrSerLeuLys 692
2792 .....GCTGTATATCTTCACTGAAGCTATCAAA 2821
693 GluValAlaProAsn.....AlaAspLeuLysAlaLeuLeuLys 706
2822 CCAGTTGCTCCAAACACACAGCAGCTCCAACTCTGCGCATTAACATT 2871
706 nIleLeuSerValAspGlyValProAlaThrThrAlaLysAlaThrHis 723
2872 AACACACAGCAGCTGGTGGTTAGTTGATTAAACAGTCAACTACACTT 2921
723 eAsnValGluPheValSerAlaAspHisValAla..... 736
2922 TAGCAATTTCATTAGCTGATCCAGATCTTAATGTAAGTCAACACTGT 2971
737 GluAsnGlyThrValGlyAlaLysGlyAlaThrSerIleThrValLysAs 753
2972 GATACTGCAACTGTTTCATTAAAGATGCAATATATCATTA...TC 3018
753 nLeuThrValValLysAspGlyLysGluGlnLysValGluPheAspLys 770
3019 TCTTACATTAGTTGAACACTGCTGCTAATACAGGTGA...TTTCTCAAA 3065
770 lValGln.....ValAlaVal 775
3066 CTGTTCACGCTGATCATTAATCTTTAACTGCTGATCATTAACAGTT 3115
776 SerIleLysGluAlaLysProAla 783
3116 ACTTATGCAAGATGCTAAATAATGCT 3139
seq_name: /stds/gcgdata/geneseq/geneseq/NA1997.DAT.AAT75487
seq_documentation_block:
ID AAT75487 standard; DNA; 3687 BP.
XX
AC AAT75487;
XX
DT 27-FEB-1998 (first entry)
XX
DE DNA for Bacillus stearothermophilus S-layer protein sbs-A.
XX
KW S-layer; sbs-A; vaccine; adjuvant; carrier; hybridisation assay;
KW molecular spinning nozzle; molecular laser; ds.
XX
OS Bacillus stearothermophilus.
XX
FH key Location/Qualifiers
FT CDS 1..3687
FT sig_peptide 1..90 /*tag= a
FT mat_peptide 91..3684 /*tag= b
FT /*tag= c
XX
PN DE19603649-A1.
XX
PD 07-AUG-1997.
XX
PF 01-FEB-1996; 96DE-1003649.
XX
PR 01-FEB-1996; 96DE-1003649.
XX

```

```

PA (LUBI/) LUBITZ W.
PA (SLEY/) SLEYTR U.
XX
PI Kuen B, Lubitz W, Sleytr U;
XX
DR WPI; 1997-394558/37.
XX
PT Preparation of S-layer proteins by expressing sbs-A gene in Gram
PT negative bacterium - or new sbs-B gene in any host, also new
PT recombinant proteins containing heterologous inserts, e.g.
XX epitoes(s), useful as vaccines and adjuvants
XX
PS Claim 1; Pages 9-14; 31pp; German.
XX
CC The present sequence encodes the Bacillus stearothermophilus PV72
CC S-layer protein, sbs-A. S-layer structures can be used as vaccines
CC or adjuvants, particularly when they include a bacterial ghost that
CC may contain additional epitopes in its membrane. Other uses of
CC recombinant sbs-A, depending on the nature of the inserted peptide,
CC are as an universal carrier for biotinylated reactants for use in
CC immunological or hybridisation assays (the insert is streptavidin),
CC to induce immune responses (epitopes), as a reagent for removing
CC cytokine or toxin from serum (antigenic epitopes), as a molecular
CC spinning nozzle (polymethoxybutyrate synthase) and as a molecular
XX laser (luciferase).
XX
SQ Sequence 3687 BP; 1316 A; 660 C; 715 G; 996 T; 0 other;
XX
alignment_scores:
Quality: 248.50 Length: 898
Ratio: 0.597 Gaps: 44
Percent Similarity: 46.325 Percent Identity: 22.383
XX
alignment_block:
US-09-754-947-1 x AAT75487 ..
XX
Align seq 1/1 to: AAT75487 from: 1 to: 3687
XX
55 AsnLeuProIleAspLysAspAlaLysProSerPheAlaAspSerGlnG 71
967 GATCTTCTTTATCTACAGACGAGAAACATACCTGATGCTGCTTA.. 1014
71 yGlnTPTrThrProPheIleAlaValGluLysAlaGlyValIlel 88
1015 .....ACTCATTCGAAATAATACGAGATATAAGTACTACTTA 1054
88 ySGlyThrGlyAsnGlyPheGluProAsnGlyLysIleAspHisValSer 104
1055 AAGGTATTAA.....GACAAAATGCGCAAGAAATTAAAGAAAGAT 1095
105 MetLaserLeuLeuValGluAlaThrLysLeuAspThrLysValAsnG 121
1096 GCATTCACTTTCACAGCTTCGAAATGATGCTGCTACTTCAAGTCTTG 1145
121 yThrProAlaThr..... 125
1146 AACTAATGTAACAACAACACTTCTGTAACTAGACAGCAAGTACTTGC 1195
126 .....LysPheLysAspLeuGlu 131
1196 ACACGTGACGATCTTTAAACAGTACTATTGATTAAGTTGATGACACCTGAA 1245
132 ThrLeuAsn.....TrpGlyLys.. 137
1246 ACTGTAAACAGCTGCAAGCTTACTATTACAGATGTGAAGTGAAGAACG 1295
138 .....GluLysAlaAsnIleLeuValGluLeuG 147
1296 CATTCAGATTAATGCACTACTTCTGCTTACAAATTAATTAATGCTTAA 1345
147 LylIleSerValGlyThrGlyAspGlnTrpGlu.....ProLys 159

```

```
1346 AAGAACGCTTACTGTAACCTGTAACCAATATTAACCTGTAATCAATATGCTT 1395
160 LysThrValThr.....LysAlaGluAlaAlaGlnPheIleAlaLysThr 174
|||||.....:|||||
1396 AAACACTTAACGTGTTACAAATGACAGAGCTTACGAGTTAGTC..... 1437
174 rAspLysGlnPheGlyThrGluAlaAlaLysValGlnSerAlaLysAlaV 191
|||||.....:|||||
1438 .....TTCACTGCACCAACGATCAGCACCAACTGTTGCTACCGCTC 1477
191 aLThrThr..... 193
|||||
1478 CTACTACTTTAGAGTGGTACAACTTATCTAGTCTTCTTCAACAAT 1527
193 ..... 193
1528 GTTTGGGGAATTAATGGCTGTGTGTAATGAAGTGAAGCTGAACCTTATTATCC 1577
194 ...GlnLysValGluValLysPheSerLysAlaValGluLysLeuThrL 209
|||||.....:|||||
1578 TGGCTCTTCAATTCAACACACCTTTGCTACTAAGTTAGACGAATCTACTT 1627
209 ySGLuAspLLeuLysValThrAsnLysAlaAsnAsnAspLysValLeuVal 225
|||||.....:|||||
1628 TAGCTGATTAACCTTTGTATTAGTTGAAGAAATCTGCTACAGTTGTTGCT 1677
226 LysGluValThrLeuSerGluAspLysArgSerAlaThrValGluLeuTy 242
|||||.....:|||||
1678 TCTGAACCTAAATATATATGACAGCCCTAAATATGTAACCTTATAGTCCAAA 1727
242 rSerAsnLeuAlaAlaLysGlnThrTyThrValAspValAsnLysValG 259
|||||.....:|||||
1728 ACCGGACCTTAAAGAAAATATCAATCTATCAAAATTAATAAAAAAGCT 1777
259 LysThrGlu.....ValAlaValGlySerLeuGluAlaLysThrIle 273
|||||.....:|||||
1778 TGAAGTCCCATTAAGGTATTTGAATTAGGACACTGTTAACAGAAAACATTT 1827
274 GluMetAlaAspGlnThrValAlaAlaAspGluProThrAlaLeuGlnPh 290
|||||.....:|||||
1828 GAGTTCAAAACCTCAAGACTTAACCTCT.....CCTACAGTTATTAAGCT 1871
290 eThrValLysAsp...GluAsnGlyThrGluValAlaSerProGluGlyI 306
|||||.....:|||||
1872 AACGCTTAAATAATGGCGAGCTGTGATTAAGTAACCT.....GAAGCTC 1915
306 LeuGlnPheValThrProAlaAlaGluLysIleAsnAlaLysGlyLeuIle 322
|||||.....:|||||
1916 AAGAAATTACTGTGAAGTTCTCAGAGAAATTTAAATATACATTTAATGCTACA 1965
323 ThrLeuAlaLysGlyThrSerThr..... 330
|||||.....:|||||
1966 ACCGTTTCGGGTAGCACATCAATACGTCAGTCAAGTTGCTGTAGTAAAGC 2015
331 .....ThrV 332
2016 GGGTCAAACTTATCTGCTCTTACAGCAAGTGACATTCATTCACGCTAGG 2065
332 aLysAlaValTyLysLysAspGly.....LysValAlaAla 344
|||||.....:|||||
2066 TTGAACCGCTTACTGTGTCAAGATGGAACATCAAAAGTGAAAGTTGCTGCT 2115
345 GluSerLysGlu..... 348
|||||
2116 AACCAATTAGAACGTAAACCAAGGTACAAATTAAGTGTTCGCTAAAGG 2165
349 .....ValLysValSerAlaGlnGlyAlaAlaValAlaSerI 361
|||||.....:|||||
2166 TGCACACACTCTCTTTAAAGATGCGCAAAATGCAAAATCTTACGCACTA 2215
361 LeSerAsnTyrThrValAlaGlnGlnAsnLysAlaAspPheThrSerLys 377
|||||.....:|||||
2216 ACTATATCTATACATTTTACAACTGAAGGTCAA...GACGTAAACGACCA 2262
378 AspPheLysGlnAsnAsnLysValTyArgLysAspAsnAlaTyValGI 394
|||||.....:|||||
2263 ACGGTT.....ACAAAGTATTCAAAGCTGATTCCT..... 2292
394 nValGluLeuLysAspGlnPheAsnAlaValThrThrGlyLysValGluT 411
|||||.....:|||||
2293 .....TTAAAGACGCT...GATGCAGTTACTACA..... 2319
411 TyrGluSerLeuAsnThrGluValAlaValAlaAspLysAlaThrGlyLys 427
2319 ..... 2319
428 ValThrValLeuSerAlaGlyLysAlaProValLysValThrVal..... 442
|||||.....:|||||
2320 CTTACGAACGTTGATGACAGCTCAA.....AAATTCATCATCCAAAT 2360
443 .....LysAspSerLysGlyLysAlaLeuValSerHisTyrV 455
|||||.....:|||||
2361 TAGCGAAGAATTAATAAACTTCTAGTGCT...TCTTATGCTGGGTGCAAG 2407
455 aGluIleGluAlaPheAlaGlnLysAlaMetLysAspIleLysLeuGlu 471
|||||.....:|||||
2408 TTAAGTGTGAGAAATTAACAACCAACGATGGGTAGAT.....GCT 2448
472 LysThrAsnValAlaLeuSerThrLysAspValThrAspLeuLysValLy 488
|||||.....:|||||
2449 GGTACTGGAACAACTGTATCAGTT..... 2472
488 sAlaProValLeuAspGlnTyArgLysGluPheThrAlaProValThrV 505
|||||.....:|||||
2473 .GCTCTTAAGACAGATGCCAAATGTAAAGTAAACACTGCTGTGTACAT 2521
505 aLysValLeuAspLysAspGlyLysGluLeuLysGluGlnLysLeuGlu 521
|||||.....:|||||
2522 TAACGCGCTTGACACATAACGACAAAGATGCGAAA...TTCCGCTCTGTA 2568
522 ALaLysTyThrValAsnArgGluLeuValLeuAsnAlaAlaGlyGlu.... 536
|||||.....:|||||
2569 GTAGATAAGTCTTCTACATGATGGAATTCGTGATGACCTGCTAATGTAAAT 2618
537 .....GluAlaGlyAsnTyThrValLy 544
2619 TAGGAAAAAGATATTTAATTCGTTACACACGCTGGAGACACTGTAG 2668
544 aLeuThrAlaLysSerGlyLysGluAlaLysAlaThrLeuAlaLeu 560
|||||.....:|||||
2669 CTTGCTGCAAAAGCTGCTGCTGACAAAGATGGTCAA..... 2703
561 GluLeuLysAlaProGlyAlaPheSerLysPheGluValArgGlyLeuAs 577
|||||.....:|||||
2704 .....AACGCTTCGCTGATTCCTCA.....ACAG 2729
577 pThrGluLeuAsp.....LysTyThrValThrGluGluAsnGlnLys 591
|||||.....:|||||
2730 CACTGCAATTAATACACTAAGAGCTTATTAAGTTGAATTCATGAA.... 2775
591 snAlaMetThrValSerValLeuProValAlaAspAlaAsnGlyLeuValLeu 607
|||||.....:|||||
2776 .....ACTGATTACCGGAAGTTAAACCTGAGAAACATCGTTGTT 2814
608 LysGlyAlaGluAlaAlaGluLeuLysValThrThrThrAsnLysGluGI 624
|||||.....:|||||
2815 AAAGATGCAGCAGGTAATGCGTAGCTGCTGCTAATACACCATTAAGACG 2864
624 yLysGluValAspAlaThrAspAlaGlnValThrValGlnAsnAsnSerV 641
|||||.....:|||||
2865 TTCT.....ACAAATTAATTTG 2881
641 aLLeuThrValGlyGlnGlyAlaLysAlaGlyGluThrTyThrLysValThr 657
|||||.....:|||||
2882 TATTACTCCATCTCAAGAAATTAAGAGCTGCTACAGATTACTCTGTAA 2931
```



```

658 ValValLeuAspGly.....LysLe 664
    :|:|:|
2932 ATT.....GACGGTGCAGAGATAAGTAGTAACACAAATCTCTAATA 2975
664 uileThrThrlSerPheLysValValAspThlAlaPthrAlaProthrAlaLysG 681
    |||||
2976 CATTAAT.....TCGTTCAAGACTGATATCGCGAATCCAAACG..... 3012
681 LysLeuAlaValGluPheThrSerThrlSerLeuLysGluValAlaProAsn 697
    :|:|:|
3013 .....TTATCTCAATCAGCATTGCTGACGGGTGACGATTAC 3048
698 AlaAspLeuLysAlaAlaLeuLeuAsnLleLeuSerValAspGlyValPr 714
    :|:|:|
3049 GTTGAC...CGTCTAAACAAATTAACAATTAATTCACAGCTTCAGTTCC 3095
714 o...AlaThrThrlAlaLysAlaThrAlaSerAsnValGluPheValSerA 730
    |||
3096 AAACCCAAACATTCACCTTAAGAAAGCTGACGAGAACTTCATTACTAAT 3145
730 LaasP...ThrAsnValValAlaGluAsnGlyThr.....ValGly 742
    :|:|:|
3146 ACACTTTAGTAATAGTAATATGAAATTAACATTAACAAATGTATTC 3195
743 AlaLysGlyAlaThr.....SerLetyrValLysAsnLeuThrValVa 757
    |||||
3196 CACAAGCTGTACACCTGACGAGTTTACTCAATATGAGTTAGCGTTTC 3245
757 LysAspGlyLys.....GluGluLysValGluPheAspL 769
    |||||
3246 AAAAGATTTTCAAACTGCTACTGATATTGATGATGACAACTTCATTC.... 3291
769 ysaLysAlaGlnValAlaValSerLetyrLysGluAlaLysProAla 783
    :|:|:|
3292 .....ATCACAGGTTCTGTCTACTACTGACGAAGTAAACCTGCT 3330
seq_name: /SIDS2/gcgdata/geneseq/NA1999.DAT:AAx22748
seq_documentation_block:
ID AAX22748 standard; DNA; 3687 BP.
XX
XX AC AAX22748;
XX
XX DT 24-AUG-1999 (first entry)
XX
XX DE B stearothermophilus sbsA DNA.
XX
XX KW sbsA protein: S-layer protein; Gram-negative; prokaryotic host cell;
    integration: cytoplasmic membrane; secretion; periplasmic space; toxin;
    eukaryotic host cell; vaccine; adjuvant; immunogenic epitope; luciferase;
    immunostimulant; cytokine; polyhydroxybutyrate; PHB synthase; body fluid;
    molecular laser; universal carrier molecule; monomolecular layer; ds.
XX
XX OS Bacillus stearothermophilus.
XX
XX PH key Location/Qualifiers
XX FT CDS 1..3687
XX FT /*tag= a
XX FT /product= "sbsA"
XX FT sig_peptide 1..90
XX FT /*tag= b
XX FT mat_peptide 91..3684
XX FT /*tag= c
XX
XX PN DE19732829-A1.
XX
XX PD 04-FEB-1999.
XX
XX PF 30-JUL-1997; 97DE-1032829.
XX
XX PR 30-JUL-1997; 97DE-1032829.
XX
XX PA (LUBIT/) LUBITZ W.

```

```

XX Lubitz W, Resch S;
PI
XX WPI; 1999-122189/11.
XX P-PDB; AAW93252.
XX
XX PT Producing S-layer proteins in Gram-negative bacteria or eukaryotes -
    integrated into membranes or organelles or secreted into periplasma
    or growth medium, and nucleic acid encoding S-layer proteins with
    peptide insertions, used in vaccines or for enzymatic reactions
XX
XX PS Claim 5, Page 9-14; 34pp; German.
XX
XX CC This invention describes a method for the production of a S-layer protein
    (I) which comprises (a) preparing a Gram-negative prokaryotic host cell
    transformed with nucleic acid (II) encoding (I), linked to a signal
    sequence (SS) that encodes a protein which causes at least one of (i)
    integration of (I) into the external or cytoplasmic membranes and/or
    (ii) secretion of (I) into the periplasmic space or extracellular medium,
    (b) culturing the cell to express (I) and (c) optionally recovering (I)
    from the membranes, periplasmic space and/or extracellular medium.
    Alternatively, a eukaryotic cell is used as host and then the SS, which
    is optional, promotes integration into the cytoplasmic membrane or an
    organelle and/or secretion into the extracellular medium. (I), and
    derived structures, may include a wide variety of polypeptide inserts and
    are useful as (i) vaccines or adjuvants (with immunogenic epitopes or
    CC immunostimulants inserts such as cytokines) (ii) as receptors (inserts
    'are enzymes, e.g. polyhydroxybutyrate (PHB) synthase for use as a
    CC molecular spinerette' for production of PHB or luciferase for use as
    CC molecular carrier (when combined with substrate and oxygen) and (iii) as
    universal carrier molecule (streptavidin is inserted) for use in
    CC hybridisation and immuno assays, or for selective elimination of
    CC cytokines, toxins etc. from body fluids (inserts are specific binding
    CC epitopes). In this system, heterologous (I) do not form inclusion bodies
    CC but rather monomolecular layers, and in eukaryotic cells they undergo
    CC glycosylation. This sequence encodes the Bacillus stearothermophilus
    sbsA protein which is used to illustrate the method of the invention.
XX
XX SQ Sequence 3687 BP; 1316 A; 660 C; 715 G; 996 T; 0 other:
XX
XX alignment_scores:
    Quality: 248.50 Length: 898
    Ratio: 0.597 Gaps: 44
    Percent Similarity: 46.325 Percent Identity: 22.383
XX
XX alignment_block:
    US-09-754-947-1 x AAX22748 ..
XX
XX Align seg 1/1 to: AAX22748 from: 1 to: 3687
XX
XX 55 AsnLeuProIleAspLysAspAlaLysProSerPheAlaAspSerGlnG 71
    :|:|:|
967 GATGTTCTTATCTACAGACGAAACATCTCACTGTGATGCTCA.. 1014
71 yGlnTrpTrThrProPheIleAlaValGluLysAlaGlyValIlel 86
    |||||
1015 .....ACTCCATTCGAAATAATATACGAGATTAAGTAACTAGTATA 1054
88 ysgLyThrGlyAsnGlyPheGluProAsnGlyLysIleAspArValSer 104
    |||||
1055 AAGGTATTAA.....GACAAAAATGCGCAAGAATTAAAGACAGAT 1095
105 MetAlaSerLeuLeuValGluAlaTyrLysLeuAspThrLysValAsnG 121
    :|:|:|
1096 GCATTCACCTTCAAGCTTCGAAATGATGCTGATGACTCAAGTGTGG 1145
121 yThrProAlaThr..... 125
    |||||
1146 AACTAATGTACAAACAACTCTGTAACTAGCAGCAGCTACTTTGG 1195
126 .....LysPheLysAspLeuGlu 131
    |||||

```

1196 ACACCTGACGACTTACCTTAAACAGTAGTATTGATTAGTTGTTAGCACCTGAA 1245
132 ThrLeuAsn.....TrpGlyLys.. 137
1246 ACTGTAAACGCTCGAACGTTACTATTACAGATGTTGAACAGTGAAAGG 1295
138GluLysAlaAsnIleLeuValGluLeuG 147
1296 CATTCCAGTAATGTCATCTACTCTGTTCTGCTACAAATTAATACTATTACGTTAA 1345
147 LylleSerValGlyThrGlyAspGlnTrpGlu.....ProLys 159
1346 AACAGCGTTAGTAAGTACTGCTAAACATATTAACCTCTCATCAATATGTT 1395
160 LysThrValThr.....LysAlaGluAlaIleGlnPheIleAlaLysThr 174
1396 AAACATTAATACGTGTTACAATGAGAGCTTACGAGTTAGTG..... 1437
174 rAspLysGlnPheGlyThrGluAlaAlaLysValGluSerAlaLysAlaVal 191
1438TTCACGCAACAGCATGACACCAACCTGTTGCTACGCGTC 1477
191 aThrThr..... 193
1478 CTACTACTTTAGTGTGTACACTTTATCTACTGTTCTCTTACACAAT 1527
193 193
1528 GTTTGGGCTAAATTGGCTGTGTGATGAATGAAGCTGAACCTTATTATCC 1577
194GlnLysValGluValLysPheSerLysAlaValGluLysLeuThrL 209
1578 TGGTCTTCAATTACACAACAGCTTTGCTACTAAGTTAGACGAATCTACTT 1627
209 yGluGlnLleLysValThrAsnLysAlaAsnAsnAspLysValLeuVal 225
1628 TAGCTCATTAAGTTGATTAAGTTGAAGAAAGATCTGGTACACTGTTGCT 1677
226 LysGluValThrLeuSerGluAspLysArgSerAlaThrValGluLeuThr 242
1678 TCTGAAGCTAAATATATATGACAGCGCTAAATGCTTAAGTGGCCAA 1727
242 rSerAsnLeuAlaAlaLysGlnThrTyThrValAspValAsnLysValG 259
1728 ACCGCGACTTAAAGAAATACATCTATCAATCAAAATTAAGAAAGGCT 1777
259 LylSerThrGlu.....ValAlaValIleSerLeuGlnAlaLysThrIle 273
1778 TGAAGTCCCATTAAGGTATTGAATTAGGACCTGTTACGAGAAACATAT 1827
274 GluMetAlaAspGlnThrValAlaAlaAspGluProThrAlaLeuGlnP 290
1828 GAGTTCAAAACATCAAGACTTAACCTGCT....CCTACAGTTATTAGCT 1871
290 eThrValLysAsp....GluAsnGlyThrGluValIleSerProGluGlyI 306
1872 AACGCTTAAAAATGGCGAGCTGGATTAAAGTAAGT....GAAGCTC 1915
306 lGluPheValThrProAlaIleGluLysIleAsnAlaLysGlyIle 322
1916 AAGATTATTAAGTTCAAGTTCTCAGAGAAATTTAAATTAATTAATGCTACA 1965
323 ThrLeuAlaLysGlyThrSerThr..... 330
1966 ACCGTTTCGGTAGCACATCAATACGCTCAAGTCTGCTAGTAAAGC 2015
331ThrV 332
2016 GGGTCAAACTTATCTGCTTACAGCAAGTGACATCAATCCAGCTAGTG 2065
332 aLysAlaValLysLysLysAspGly.....LysValAlaAla 344
2066 TTGAAGCGTTACTGCTCAAGATGGAACATACAAAGTGAAAGTCTGCT 2115
345 GluSerLysGlu..... 348
2116 AACCATATTGAAGCTAACCAAGGTACAAATTAAGTGTTCGGTAAAGG 2165
349ValLysValSerAlaGluGlyAlaAlaValAlaSerI 361
2166 TGCACAGCTCCTGTTAAAGATGTCGCAAAATGCAATACTTTAGCACTA 2215
361 lSerAsnTrpThrValAlaGluGlnAsnLysAlaAspPheThrSerLys 377
2216 ACTATATCTATCATTTACAACTGAAGGTCAA..GAGCTAACACACACA 2262
378 AspPheLysGluAsnAsnLysValTyrgLysAspAsnAlaTyraValG 394
2263 ACGGTT.....ACAAAGTATTCAAGGTGATTC..... 2292
394 nValGluLeuLysAspGlnPheAsnAlaValThrThrGlyLysValGluT 411
2293TTAAAGAGCGT...GATGCAATTACTTACA..... 2319
411 yrGluSerLeuAsnThrGluValAlaValAlaAspLysAlaThrGlyLys 427
2319 2319
428 ValThrValLeuSerAlaGlyLysAlaProValLysValThrVal..... 442
2320 CTTACGAAGCTTGATGCAGGTCAA.....AAATTCATATCCAAAT 2360
443LysAspSerLysGlyLysAlaLeuValSerThrIleThrV 455
2361 TAGCGAAGATTAAAACTTCTAAGTGT...TCTTATGTTGGTGGCAAG 2407
455 aGluIleGluAlaPheAlaGlnLysAlaMetLysAspIleLysLeuGlu 471
2408 TTAAGTGTGAGAAATTACAAACACGCGATGGTGAT.....GCT 2448
472 LysThrAsnValAlaLeuSerThrLysAspValThrAspLeuLysValL 488
2449 GGTACTGGAACACGTATCAAGT..... 2472
488 sAlaProValLeuAspGlnTyrgLysGluPheThrAlaProValThrV 505
2473 .GCTCTTAAGACAGATGCAATGTGTAAAGTAAACACCTGCTGTGTTAT 2521
505 aLysValLeuAspLysAspGlyLysGluLeuLysGluGlnLysLeuGlu 521
2522 TAACTGTGCTTGACAAATTAACGACAAAGATGCGAAA..TTGCGTGTGTA 2568
522 AlalysTyrrValAsnArgGluLeuValLeuAsnAlaIleGlyGln.... 536
2569 GTAGTAAAGCTTCTTACTGATGATGATTCGTGATGAGCTGTAATGTAAT 2618
537GluAlaGlyAsnTyrrThrValI 544
2619 TAAGCAAAAAGATATTTAATTCGTACACAGCTGGAGACACACTGTAG 2668
544 aLeuThrAlaLysSerGlyGluLysGluAlaLysAlaThrLeuAlaLeu 560
2669 CTTCTGTAAAGCTGCTGCTGACAAAGATGCTCAA..... 2703
561 GluLeuLysAlaProGlyAlaPheSerLysPheGluValArgGlyLeuAs 577
2704AACGCTTCGCTGATTCCTCA.....ACAG 2729
577 PThrGluLeuAsp.....LysTyrrValThrGluGlnAsnGlnLysA 591
2730 CACTGCATTTGATCACTAAGAGCTTATTAAGTTGAATTAATCAATAA... 2775
591 snAlaMetThrValSerValLeuProValAspAlaAsnGlyLeuValLeu 607
2776ACTGATTTAGCGAAGTTAAACCTGGAACATGCTGTT 2814


```
1246 ACCTGAACGCTCGAAGCTTACTATTACAGATGTTGAAAGTGGAAAACG 1295
138 .....GluysAlaasnIleValGluLeu 147
1296 CATTCCAGTAATTCATCTACTTCTGCTTCACAAATTACTATTACGTTAA 1345
147 LylSerValGlyThrGlyAspGlnTrpGlu.....ProLys 159
1346 AAGAACGCTTAGTAAGTGGTAACAAATATTAATCTGCTATCATATATGTT 1395
160 LysThrValThr.....LysAlaGluAlaAlaGlnPheIleAlaLysThr 174
1396 AAAACATTAACCTGTTACATGCGAAGCTTACGAGCTAGTG..... 1437
174 AspLysGlnPheGlyThrGluAlaAlaLysValGluSerAlaLysAlaVal 191
1438 .....TTCACGCAACGCAATCGACCACTGTTGCTACCGCTC 1477
191 alnThrThr..... 193
1478 CTACTACTTTAGGTGGTACAACTTATCTACTGTTCTCTACACAAT 1527
193 ..... 193
1528 GTTTGGGGAATATTGGCTGGTGGTGAATGAAGCTGAACCTTATTATCC 1577
194 .....GlnLysValGluValLysPheSerLysAlaValGluLysLeuThrL 209
1578 TGCTCTTCATTCACAAACAGCTTCTCTACTAAGTAGAGAACTTACTT 1627
209 yAsGluAspIleLysValThrAsnLysAlaAsnAspLysValLeuVal 225
1628 TAGCTGATAACTTGTGATTAGTTAGTGAAGAAATCTGCTACAGTGTGGCT 1677
226 LysGluValThrLeuSerGluAspLysArgSerAlaThrValGluLeuTy 242
1678 TCTGAAGCTAAATATATATATGACAGCGCTAAATGTAACTTATGTCGCAA 1727
242 rSerAsnLeuAlaAlaLysGlnThrTyThrValAspValAsnLysValG 259
1728 AGCGACCTTAAAGAAATATACATCTATCAATCAAAATTAAGGCT 1777
259 LysThrGlu.....ValAlaValGlySerLeuGluAlaLysThrIle 273
1778 TGAAGTCCCATTAAGCTATTGAATTAGCGACTGTAAACGAGAAACATAT 1827
274 GluMetAlaAspGlnThrValAlaAlaAspGluProThrAlaLeuGlnPh 290
1828 GAGTTCAAACTCAAGACTTAACCTGCT.....CCTACAGTTATTAGCGT 1871
290 eThrValLysAsp...GluAsnGlyThrGluValValSerProGluGlyI 306
1872 AAGCTTAATAAATGGCGACGCTGATTAAGTAACT.....GAAGCTC 1915
306 IeGluPheValThrProAlaAlaGluLysIleAsnAlaLysGlyLeuIle 322
1916 AAGAAATTTACTGTAAGTCTCAGAGAAATTAATACATTATTAAGCTACA 1965
323 ThrLeuAlaLysGlyThrSerThr..... 330
1966 ACCGTTTCGGGTAGCACATACATACGTCAGTTGCTGTAGTAAAGC 2015
331 .....ThrV 332
2016 GGGTGAACCTTATCTGCTTACAGCAAGTGACATCATTCAGCAGTAGG 2065
332 alLysAlaValTyLysLysAspGly.....LysValValAla 344
2066 TTGAAGCGCTTACTGCTCAAGATGCAACATCAAGTGAAGAGTCTGCTGCT 2115
345 GluSerLysGlu..... 348
```

```
2116 AACCAATTAGAACGTAAACCAAGGTACAAATTAAGTACTGTGCGTAAAGG 2165
349 .....ValLysValSerAlaGluGlyAlaAlaValAlaSerI 361
2166 TGCACAGCTCCTCTTAAGATGCTGCAATGCAATTAATCTTTAGCAACTA 2215
361 lSerAsnTrpThrValAlaGluGlnAsnLysAlaAspPheThrSerLys 377
2216 ACTATATCTATACATTTACAACTGAAGGTCAA...GAGCTAACACACACA 2262
378 AspPheLysGlnAsnAsnLysValTyTrpGluGlyAspAsnAlaTyValG 394
2263 ACGGTT.....ACAAAGTATTCAAGGTGATTC..... 2292
394 nValGluLeuLysAspGlnPheAsnAlaValThrThrGlyLysValGluTr 411
2293 .....TTAAAGACGCT...GATGCACTTACTACA..... 2319
411 yrcLysrLeuAsnThrGluValAlaValAlaAspLysAlaThrGlyLys 427
2319 ..... 2319
428 ValThrValLeuSerAlaGlyLysAlaProValLysValThrVal.... 442
2320 CTTACGAACGTTGATGACAGGTCAA.....AATTCACATCAAT 2360
443 .....LysAspSerLysGlyLysAlaLeuValSerHisThrV 455
2361 TACGAGACATTTAAAAACTCTTAGTGT...TCTTAGCGGTGGCCAAAG 2407
455 alGluIleGluAlaPheAlaGlnLysAlaMetLysAspIleLysLeuGlu 471
2408 TACGCTGCAGAAATTAACAACACGATGGCTGAT.....GCT 2448
472 LysThrAsnValAlaLeuSerThrLysAspValThrAspLeuLysValLy 488
2449 GGTACTGGAACAACGTATCACT..... 2472
488 sAlaProValLeuAspGlnTyTrpGlyLysGluPheThrAlaProValThrV 505
2473 .GCTCTTAAGACAGATGCAATAGTAAAGTAAACACTGCTGTGTTACT 2521
505 alLysValLeuAspLysAspGlyLysGluLeuLysGluGlnLysLeuGlu 521
2522 TAACTGCGCTTGACAAATTAACGAAAGATGCGAA...TTGCGTCTGCTA 2568
522 AlaLysTyrrValAsnArgGluLeuValLeuAsnAlaAlaGlyGln.... 536
2569 GTAGATTAAGCTTCTTACTGATGGAATTCGTGATGCTGTAATGTAAAT 2618
537 .....GluAlaGlyAsnTyrrThrValV 544
2619 TAAAGAAAAAGATATTTAATTGTTTACACAGCGCTGGAGACACACTGAG 2668
544 alLeuThrAlaLysSerGlyLysGluAlaLysAlaThrLeuAlaLeu 560
2669 CTTCTGTGAAGCTGCTGCTGCAAAAGATGTCGTA..... 2703
561 GluLeuLysAlaProGlyAlaPheSerLysPheGluValArgGlyLeuAs 577
2704 .....AACGCTTCTGCTCATTCCTCA.....ACAAAG 2729
577 pThrGluLeuAsp.....LysTyrrValThrGluGluAsnGlnLysA 591
2730 CACTGCATTTGATNCACTAAGAGCTTATTAGTGAATTCATGA... 2775
591 snAlaMetThrValSerValLeuProValAspAlaAsnGlyLeuValLeu 607
2776 .....ACTGATTTACGCGAAGTTAAACCTGAGAAACATGCTGTT 2814
608 LysGlyAlaGluAlaAlaGluLeuLysValThrThrThrAsnLysGluGl 624
2815 AAAGATGCAAGCAATATGCGGTAGCTGCTGCTGTAACAGCATTAAGACGG 2864
```


126 LysPheLysAspLeuGlu 131
 1166 ACACGACGATCTTAAACAGTAGTATTGATAGAGTTGTACGACCTGAA 1215
 132 ThrLeuAsn.....TrpGlyLys.. 137
 1216 ACTGTAACAGCGTCGAACGCTACTATTACAGATGTTGAACAGTGAACG 1265
 138 GluLysAlaAsnIleLeuValGluLeuG 147
 1266 CATTCAGTAATGTCATCTACTTCTGCTTCAATTAATTACTATTACGTTAA 1315
 147 LylSerValGlyThrGlyAspGlnTrpGlu.....ProLys 159
 1316 AAGAACGCTTAGTAAGTAAACATATAACTTCCTATCAATATATGTT 1365
 160 LysThrValThr.....LysAlaGluAlaIleGlnPheIleAlaLysTh 174
 1366 AAACATTAACGTGTTACAAATGCAGAACCTTAGAGTTAGTG..... 1407
 174 RasPlySGlnPheGlyThrGluAlaAlaLysValGluSerAlaLysAlaV 191
 1408TTCACTGCAAGCGATCAGCACCACTGTTGCTACCGCTC 1447
 191 AlThrThr..... 193
 1448 CTACTACTTTAGGTGTGACACTTTATCTACTGTTCTCTTACACAANA 1497
 193 193
 1498 GTTGGGCGTAATGCGCTGCTGTGTGAATGAAGCGAAGCTTATTATCC 1547
 194GlnLysValGluValLysPheSerLysAlaValGluLysLeuThr 209
 1548 TGGTCTCTCAATTCACAACAACGTTTCTACTAGTAGACGAATCTACTT 1597
 209 ysgLysPheLysValThrAsnLysAlaAsnAsnAspLysValLeuVal 225
 1598 TAGCTGATTAACCTTGTATTAGTTGAAGAAAGAAATCTGTACAGTTGTGCT 1647
 226 LysGluValThrLeuSerLysPlySargSerAlaThrValGluLeuTy 242
 1648 TCTGAACCTAAATATATATGACAGCGCTAAATGTAACTTGTAGTGCAAA 1697
 242 rSerAsnLeuAlaAlaLysGlnThrTyThrValAspValAsnLysValG 259
 1698 ACGGACCTTAAAGAAATACATCATCAATCAAAATTTAAAGAGCT 1747
 259 LylSerThrGlu.....ValAlaValGlySerLeuGluAlaLysThrIle 273
 1748 TCGAAGTCCGATTAAGGTATTGATTAGCGACGTGTAAACGAAACATAT 1797
 274 GluMetAlaAspGlnThrValAlaAlaAspGluProThrAlaLeuGlnh 290
 1798 GAGTTCAAACCTCAAGACTTAAGTCT.....CCTACAGTTATTAGCGT 1841
 290 eThrValLysAsp...GluAsnGlyThrGluValLysSerProGluGlyI 306
 1842 AACGCTTAATAATGGCGACGCTGCATTTAAAGTAACT.....GAAGCTC 1885
 306 LeGluPheValThrProAlaAlaGluLysIleAsnAlaLysGlyGluIle 322
 1886 AAGAAATTTACTGTAAGTCTCAGAGAAATTTAAATACATTAAAGCTACA 1935
 323 ThrLeuAlaLysGlyThrSerThr..... 330
 1936 ACCGTTTCGGTAGCACATACATACGCTCAAGTTCGTAGTAAAGC 1985
 331ThrV 332
 1986 GGGTGCAAACTTATCTGCTTACAGCAAGTACATCTCCAGCTAGTG 2035

332 AlLysAlaValTyrrLysLysAspGly.....LysValValAla 344
 2036 TTGAAGCGGCTTACTGCTCAAGATGGACATACAAAGTGAAGTCTGCTGT 2085
 345 GluSerLysGlu..... 348
 2086 AACCAATATTGAACGTAAACCAAGGTACAAATTAAGTGTGCTGAAGC 2135
 349ValLysValSerIleGluGluAlaValAlaValAlaSerI 361
 2136 TGCACAGCTCCTGTTAAAGATGCTGCNAATGCAATACTTAAAGACATA 2185
 361 LeSerAsnTrpThrValAlaGluGlnAsnLysAlaAspPheThrSerLys 377
 2186 ACTATATCATATCAATTTACAACTGAAGGTCAA...GACGTACACACCA 2232
 378 AspPheLysGlnAsnAsnLysValTyrrGluGlyAspAsnAlaTyrrValG 394
 2233 ACGGTT.....ACAAAGTATTCAAAAGCTGATTTCT..... 2262
 394 nValGluLeuLysAspGlnPheAsnAlaValThrThrGlyLysValGluT 411
 2263TTAAAGACGCT..GATGCACTTACTTACA..... 2289
 411 YrGluSerLeuAsnThrGluValAlaValAlaValAspLysAlaThrGlyLys 427
 2289 2289
 428 ValThrValLeuSerAlaGlyLysAlaProValLysValThrVal..... 442
 2290 CTTCAGAACGTTGATGACAGGTCAA.....AAATTCACATATCAATT 2330
 443LysAspSerLysGlyLysAlaLeuValSerHisThrV 455
 2331 TAGCGAAGAAATTAATACTTCTAGTGCT...TCTTATGCGGTGCCAAG 2377
 455 aLgLuIleGluAlaPheAlaGluLysAlaMetLysAspIleLysLeuGlu 471
 2378 TAACTGTCGCAAAATTAAACAAACACGATGCGTGAT.....GCT 2418
 472 LysThrAsnValAlaLeuSerThrLysAspValThrAspLeuLysValL 488
 2419 GGTACTGGAAACATGATACAGT..... 2442
 488 sAlaProValLeuAspGlnTyrrGlyLysGluPheThrAlaProValThrV 505
 2443 .GCTCTTAAGACAGATGCAATGTAAGTAAAGTAAACACTGCTGCTGTTACAT 2491
 505 AlLysValLeuAspLysAspGlyLysGluLeuLysGluGlnLysLeuGlu 521
 2492 TAACTGTCCTTGACAAATTAACGCAAAAGATGCGAAA...TTGCGTCTGTA 2538
 522 AlAlaTyrrValAsnArgGluLeuValLeuAsnAlaIleGln..... 536
 2539 GTAGATTAAGTCTTCTACTGATGGAATTCCTGATGAGCTGGAATTAAT 2588
 537GluAlaGlyAsnTyrrThrVal 544
 2589 TAAAGAAAAGATATTTAATTGTTTACACACGCTGAGACACACTGTAG 2638
 544 AlLeuThrAlaLysSerGlyLysGluAlaLysAlaThrLeuAlaLeu 560
 2639 CTTCTGTGAAGCTGCTGCTGCAAAAGATGTCGCA..... 2673
 561 GluLeuLysAlaProGlyAlaPheSerLysPheGluValArgGlyLeuAs 577
 2674AACGCTTCTGTCGATTCACA.....ACAAG 2699
 577 PThrGluLeuAsp.....LysTyrrValThrGluGluAsnGlnLys 591
 2700 CACTGCAATGTATACAACTAAGAGCTTATTTAGTTAAATTCATGAA.... 2745
 591 snAlaMetThrValSerValLeuProValAspAlaAsnGlyLeuValLeu 607

```

2746 .....ACTGATTACCGGAGATTAAACCTGAGACATCGTTC 2784
608 LysGlyAlaGluAlaGluLeuLysValThrThrAsnLysGluG1 624
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2785 AAGATGACAGAGTAATGCGTAGCTGCTGATGACAGATTGACAGG 2834
624 yLysGluValAspAlaThrAspAlaGluValThrValGlnAsnSery 641
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2835 TTCT.....ACAATAAATTG 2851
641 allIethrValGlyGlnGlyAlaLysAlaGlyLthrThrLysValThr 657
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2852 TATTCACCTCCATCTCAGAGATTAAAGCTGGTACGTTACTGTATACA 2901
658 ValValLeuAspGly.....LysLe 664
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2902 ATT.....GACGGTGCAGAGATAAGTAGGTACACATCTCAATA 2945
664 uIleThrThrIserPheLysValValAspThrAlaProthrAlaLysG 681
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2946 CATTAAT.....TCGTTCAAGACTGTATCTGCGAATCCACG..... 2982
681 LylLeuAlaValGluPheThrSerThrSerLeuLysGluValAlaProsn 697
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2983 .....TTATCTTCATCAGCATTCGTCGACGGTGCAGTTATAC 3018
698 AlaAspLeuLysAlaAlaLeuLeuAsnIleLeuSerValAspGlyValPr 714
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3019 GTTGAC...CGTTCTAAACAATTCATTCATTCATTCAGCATTCGTTCC 3065
714 O...AlaThrThrAlaLysAlaThrAlaSerAsnValGluPheValSera 730
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3066 AAACCCACCAATCACTCTTAAGAAAGCTGACGACTTCATTACTAATT 3115
730 laasp...ThrAsnValValAlaGluAsnGlyThr.....ValGly 742
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3116 ACACCTTAGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATC 3165
743 AlaLysGlyAlaThr.....SerIleThrValLysAsnLeuThrValVa 757
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3166 CACAAGGTGAACACTGACGAGTTTACTCAATGATGATGACGATTTC 3215
757 LysAspGlyLys.....GluGlnLysValGluPheAspL 769
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3216 AAAAGATTTCAAACTGCTACTGATGATGATGATGATGATGATGATGATC 3261
769 ysaIaValGlnValAlaValSerIleLysGluAlaLysProAla 783
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3262 .....ATCACAAGTTCTGTGCTACTGACGAGTAACCTGCT 3300
seq_name: /SID52/gcgdata/geneseq/geneseq0/NA1996.DAT: AAT08695
seq_documentation_block:
ID AAT08695 standard; DNA: 3832 BP.
XX
AC AAT08695:
XX
DT 15-JUL-1996 (first entry)
XX
DE sbSA gene encoding S-layer protein.
XX
KW sbSA gene; S-layer; cell surface layer; expression; ss
XX
OS Bacillus stearothermophilus.
XX
FH Key location/Qualifiers
FT CDS 37..3723
FT RBS /*tag- a
FT /*tag- a
FT /*tag- b
FT 5'UTR 1..36
FT /*tag- c

```

```

FT 3'UTR 3723..3832
FT sig-peptide /*tag- d
FT /*tag- e
FT /*tag- e
FT mat-peptide /note= "claim 1"
FT /*tag- f
FT stem_loop /note= "claim 3"
FT /*tag- g
FT /*tag- g
FT /note= "terminator"
XX
XX DE4425527-AL.
XX
XX 25-JAN-1996.
XX
XX 19-JUL-1994; 94DE-4425527.
XX
XX 19-JUL-1994; 94DE-4425527.
XX
XX (VOGE-) VOGELBUSCH GMBH.
XX
XX Lubitz W.
XX
XX WPI, 1996-077933/09.
XX
XX P-PDSB; AAR7673.
XX
XX Nucleic acid encoding signal peptide of Bacillus stearothermophilus
XX S-layer protein - which has a lysine content of at least 10 per
XX cent.
XX
XX Disclosure; Page 6-7; 12pp; German.
XX
XX The claimed signal peptide encoding sequence (see tag e of AAT08695)
XX is pref. operably linked at the 3' end to a protein encoding
XX sequence. The protein is pref. the S-layer protein (see tag f of
XX AAT08695). At the 5' end, the signal peptide encoding sequence is
XX pref. linked to an expression control sequence, pref. the sequence
XX given in AAT08696.
XX
XX Sequence 3832 BP; 1355 A; 686 C; 749 G; 1042 T; 0 other;
XX
alignment_scores:
Quality: 248.50 Length: 898
Ratio: 0.597 Gaps: 44
Percent Similarity: 46.325 Percent Identity: 22.383
alignment_block:
US-09-754-947-1 x AAT08695 ..
Align seg 1/1 to: AAT08695 from: 1 to: 3832
55 AsnLeuProIleAspLysAspAlaLysProSerPheAlaAspSerGlnG1 71
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1003 GATGTTCTTATATCTACAGCGAATAATCTACCTGATGCTCA.. 1050
71 yGlnIleThrThrProPheIleAlaAlaValGluLysAlaGlyValIleL 88
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1051 .....ACTCCATTCGAAATAATATACGAGATGATTAAGTAGACTTA 1090
88 ySGlyThrGlyAsnGlyPheGluProAsnGlyLysIleAspArgValSer 104
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1091 AAGGTATTAA.....GACAAAAATGCGCAAGAAATTAAAGAGAGAT 1131
105 MetAlaSerLeuLeuValGluAlaThrLysLeuAspThrLysValAsnG1 121
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1132 GCATTCACCTTCACAGCTTGAAATGATGCTGTAGTACTCAAGTGTGG 1181
121 yThrProAlaThr.....
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1182 AACTAATGTAAACAACAACACTTCTGTAACTTAGCAGCAGTACTTTCG 1231

```

```

126 .....LysPheLysAspLeuGlu 131
1232 ACACGTACGATCTTAAACAGTAGATTGATTGATTAGTACACTGAA 1281
1332 ThrLeuAsn.....TyrGlyLys.. 137
1282 ACTGTAAACAGCTCGAACGTTACTATTACAGATGTTGAAACCTGGAACG 1331
138 .....GluLysAlaAsnIleLeuValGluLeuG 147
1332 CATTCCAGTAATTGCATCTACTCTCTGCTTCAATTAATTACTATTACGTTAA 1381
147 LYLSErValGlyThrGlyAspGlnTyrGlu.....ProLys 159
1382 AAGAACGCTTAGTAAGTGGTAAACAATATAACTCTATCATATAATGTT 1431
160 LysThrValThr.....LysAlaGluAlaAlaGlnPheIleAlaLysTh 174
1432 AAACACTTAATGCTGTACATAGCAGAACTTACGAGTATGTG..... 1473
174 rAspLysGlnPheGlyThrGluAlaAlaLysValGluSerAlaLysAlaV 191
1474 .....TTCACGTCAAACGACATCAGACCAACTGTTGCTACGCGTC 1513
191 aLThrThr..... 193
1514 CTACTACTTAAAGTGTACAACTTATCTACTGTTCTCTTACAAACAAT 1563
193 ..... 193
1564 GTTTGGGGTAATTGGCTGTGTCGATGAAGCTGAAGCTTAATTATTC 1613
194 ...GlnLysValGluValLysPheSerLysAlaValGluLysLeuThrL 209
1614 TGCTCTTCAATTCACAAACAAGCTTGTCTACTAAGTTAGACAGATCATCTT 1663
209 yAspLysAlaLysValThrAsnLysAlaAsnAsnAspLysValLeuVal 225
1664 TAGCTGATTAACCTTGTATTAGTTGAAAAAGAAATCTGTACAGCTGTGTGCT 1713
226 LysGluValThrLeuSerGluAspLysArgSerAlaThrValGluLeuTyr 242
1714 TCTGAACCTAAATATATATGACAGCGCTAAATATGTTACTTATAGTGCAAA 1763
242 rSerAsnLeuAlaAlaLysGlnThrTyrThrValAspValAsnLysValG 259
1764 ACCGCGACCTTAAAGAAATFCAATCATCAATCAAAATTAATAAAAGGCT 1813
259 LysThrGlu.....ValAlaValGlySerLeuGluAlaLysThrIle 273
1814 TGAAGTCCGATTAAGGTATTGAATTAGCGACCTGTAAACGAAACATAT 1863
274 GluMetAlaAspGlnThrValAlaAlaAspGluProThrAlaLeuGlnPh 290
1864 GAGTTCAAAACATCAGACTTAACCTCT.....CTACAGTTATTAGCGT 1907
290 eThrValLysAsp...GluAsnGlyThrGluValAlaSerProGluGlyI 306
1908 AACGTCCTAAATAATGCGACGCTGATTTAAAGTAACT.....GAAGCTC 1951
306 LeuGluPheValThrProAlaAlaGluLysIleAsnAlaLysGlyGluIle 322
1952 AAGAAATTTACTGTGAAGTTCTCAGAGAAATTTAAATACATTAAATGCTACA 2001
323 ThrLeuAlaLysGlyThrSerThr..... 330
2002 ACCGTTTGGGTAGCACATCATACGTCAAGTTGCTGTAGTAAAGC 2051
331 .....ThrV 332
2052 GGGTGAACAACTTATCTGCTTACAGCAAGTCACATCATTCACAGCTAGTG 2101
332 aLysAlaValThrLysLysAspGly.....LysValValAla 344

```

```

2102 TTGAAGCGGTACTGTCGTCAAAGATGGAACATACAAAGTGAAGTCTGCT 2151
345 GluSerLysGlu..... 348
2152 AACCAATTAGAAAGTAAACGTAACCAAGGTACAAATTAGTAGTTCGGTAAAG 2201
349 .....ValLysValSerAlaGluGluAlaAlaValAlaLaserI 361
2202 TGCACACGTCCTCTGTTAAAGATGCTGCACAAATGCAAATTAATTAGCACTA 2251
361 LeSerAsnTyrThrValAlaAlaGluGlnAsnLysAlaAspPheThrSerLys 377
2252 ACTATATATATCATATTACAACTGAAGGTCA...GACGTACACACACCA 2298
378 AspPheLysGlnAsnAsnLysValTyrGluGluLysAspAsnAlaTyrValG 394
2299 ACGGTT.....ACAAAGTATTAAGGTGATTC..... 2328
394 nValGluLeuLysAspGlnPheAsnAlaValThrThrGlyLysValGluT 411
2329 .....TTAAAGACGCT..GATCGAGTTACTACA..... 2355
411 TyrGluSerLeuAsnThrGluValAlaValAlaValAspLysAlaThrGlyLys 427
2355 ..... 2355
428 ValThrValLeuSerAlaGlyLysAlaProValLysValThrVal..... 442
2356 CTTACGAACTGTGATGACAGCTCA.....AAATTCACATCATCATT 2396
443 .....LysAspSerLysGlyLysAlaLeuValSerHisThrV 455
2397 TAGCGAAGATTAATAAACTTCTAGTGT...TCTTATGCGGTGCAAG 2443
455 aLGIuIleGluAlaPheAlaGlnLysAlaMetLysAspIleLysLeuGlu 471
2444 TAAGTGTGAGAAATTAACAACAACAGAGATGGTAT.....GCT 2484
472 LysThrAsnValAlaLeuSerThrLysAspValThrAspLeuLysValLys 488
2485 GGTACTGGACACACGTATCACTT..... 2508
488 sAlaProValLeuAspGlnTyrGlyLysGluPheThrAlaProValThrV 505
2509 ..GCTCTTAAGACAGATGCCAAATGTAAAGTAAACGCTGCTGTGTTACAT 2557
505 aLlysValLeuAspLysAspGlyLysGluLeuLysGluGlnLysLeuGlu 521
2558 TAACTGCTTGTACATTAACGCAAAAGATGCGAAA...TTCGCTGTGTA 2604
522 AlaLysTyrValAsnArgLyuLeuValLeuAsnAlaAlaGlyGln..... 536
2605 GTAGATTAAGTCTTCTACTGATGGAATTCGTGATGAGCTGTAATGTAAT 2654
537 .....GluAlaGlyAsnTyrThrValV 544
2655 TAAGAAAAAGATATTTAATTCGTTTACACAGCTGCGACACACTGTAG 2704
544 aLLeuThrAlaLysSerGlyGluLysGluAlaLysAlaThrLeuAlaLeu 560
2705 CTTCTGTGAAGCTGCTGCTGACAAAGATGGTCA..... 2739
561 GluLeuLysAlaProGlyAlaAlaPheSerLysPheGluValArgGlyLeuAs 577
2740 .....AACGCTTCTGCTGCAATTCCA.....ACAG 2765
577 ThrGluLeuAsp.....LysTyrValThrGluGluLysAsnGlnLysA 591
2766 CACTGCATTTGATACAACTAAGAGCTTATTAGTTGAATTCAATGA... 2811
591 snAlaMetThrValSerValLeuProValAspAlaAsnGlyLeuValLeu 607

```



```

2812 .....ACTGATTACGGAGTTAAAGCTGAGCAATCGTTGTT 2850
608 LysGIAlaGIuAlaAlaGIuLeuLysValThrThrThrAsnLysGIuGI 624
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2851 AAAGATGACGACGAGTAATGCGGTGAGCTGACTGACGTACACGATTAACG 2900
624 LysLysIuValAspAlaThrAspAlaGIuValThrValGIuAsnAspSer 641
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2901 TTCT.....ACAAATAAATTTTG 2917
641 allIethrValGIuGIuGIuLysAlaGIuLysAlaGIuThrThrLysValThr 657
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2918 TATTCACCTCCATCTCCAGAAATTAACCTGGTACAGTTTCTCTGTAAACA 2967
658 ValValLeuAspGly.....LysIle 664
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2968 ATT.....GACGGTGCAGAGATAAGTAAAGTAAACATCTCTAAATA 3011
664 uIleThrThrHisSerPheLysValValAspThrAlaProThrAlaLysG 681
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3012 CATTACT.....TCGTTCAAGACTGTATCTGCGAAACCAACG..... 3048
681 LysLeuAlaValGIuPheThrSerThrSerLeuLysGIuValAlaProAsn 697
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3049 .....TTATCTTCAATCAGCATTTGCTGACGGTGCAGTTAAC 3084
698 AlaAspLeuLysAlaAlaLeuLeuAsnIleLeuSerValAspLysValPr 714
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3085 GTTGAC.....CGTTCAAAACAATTACAAATTGAATTCAGCATTTCACTGC 3131
714 C...AlaThrThrAlaLysAlaThrAlaSerAsnValGIuPheValSerA 730
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3132 AAACCCAAACAATCACTCTTAAGAAAGCGTGACGCAATCTTACAAAT 3181
730 laasp...ThrAsnValValAlaGIuLysnGIuThr.....ValGIy 742
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3182 ACACCTTGTAAATGTAAATATATATAAATAAACAATCAAAATGTATTTC 3231
743 AlaLysGIuAlaThr.....SerIleTyValLysAsnLeuThrValVa 757
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3232 CACAAAGTGTAAACCTTGACGAGTTTACCAATATGACGTAGACAGTTTC 3281
757 LysAspGIuLys.....GIuGIuLysValGIuPheAspL 769
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3282 AAAACATTTTCAACCTGCTACTGATATTGATAGCAAGATTACATTC.... 3327
769 ysAlaValGIuValAlaValSerIleLysGIuAlaIleProAla 783
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3328 .....ATCACAGGTTCTGTGTCTACTGACGAGAAATAAACCTGCT 3366
seq_name: /SID2/gcgdata/geneseq/geneseqn/MA1999.DAT:AAx78245
seq_documentation_block:
ID AAX78245 standard; DNA: 4988 bp.
XX
XX AAX78245:
XX
XX
XX 24-AUG-1999 (first entry)
XX
XX
XX B. steaerothermophilus sbSA/male fusion protein DNA sequence.
DE
XX sbSA protein: S-layer protein; Gram-negative; prokaryotic host cell;
XX integration; cytoplasmic membrane; secretion; periplasmic space; toxin;
XX eukaryotic host cell; vaccine; adjuvant; immunogenic epitope; luciferase;
XX immunostimulant; cytokine; polyhydroxybutyrate; PHB synthase; body fluid;
XX molecular laser; universal carrier molecule; monomolecular layer;
XX fusion protein; male; ss.
XX
XX
XX Synthetic.
XX
XX Bacillus steaerothermophilus.
XX
XX DE19732829-AL.
XX
XX

```

04-FEB-1999.
XX
PF 30-JUL-1997: 97DE-1032829.
XX
XX 30-JUL-1997: 97DE-1032829.
PA
XX (LUBITZ/) LUBITZ W.
PI
XX Lubitz W., Resch S.
DR WPJ: 1999-122189/11.
XX
XX
PT Producing S-layer proteins in Gram-negative bacteria or eukaryotes -
PT integrated into membranes or organelles or secreted into periplasma
PT or growth medium, and nucleic acid encoding S-layer proteins with
PT peptide insertions, used in vaccines or for enzymatic reactions
XX
PS
XX Example 7; Fig 4; 34pp: German.

This invention describes a method for the production of a S-layer protein
CC (I) which comprises (a) preparing a Gram-negative prokaryotic host cell
CC transformed with nucleic acid (II) encoding (I), linked to a signal
CC sequence (SS) that encodes a protein which causes at least one of (i)
CC integration of (I) into the external or cytoplasmic membranes and/or
CC (ii) secretion of (I) into the periplasmic space or extracellular medium
CC (b) culturing the cell to express (I) and (c) optionally recovering (I)
CC from the membranes, periplasmic space and/or extracellular medium.
CC Alternatively, a eukaryotic cell is used as host and then the SS, which
CC is optional, promotes integration into the cytoplasmic membrane or an
CC organelle and/or secretion into the extracellular medium. (I), and
CC derived structures, may include a wide variety of polypeptide inserts and
CC are useful as (i) vaccines or adjuvants (with immunogenic epitopes or
CC immunostimulants) inserts such as cytokines) (ii) as reactors (inserts
CC are enzymes, e.g. polyhydroxybutyrate (PHB) synthase for use as a
CC 'molecular spinnerette' for production of PHB or luciferase for use as
CC molecular laser (when combined with substrate and oxygen) and (iii) as
CC universal carrier molecule (streptavidin is inserted) for use in
CC hybridisation and immuno assays, or for selective elmination of
CC cytokines, toxins etc. from body fluids (inserts are specific binding
CC epitopes). In this system, heterologous (I) do not form inclusion bodies
CC but rather monomolecular layers, and in eukaryotic cells they undergo
CC glycosylation. This sequence encodes a novel fusion protein constructed
CC from the Bacillus sterothermophilus sbxa gene and the mle gene signal
CC sequence which is used to illustrate the method of the invention.
XX
SQ Sequence 4988 BP; 1678 A; 985 C; 1052 G; 1273 T; 0 other;

alignment_scores:
Quality: 248.50 Length: 898
Ratio: 0.597 Gaps: 44
Percent Similarity: 46.325 Percent Identity: 22.383

alignment_block:
US-09-754-947-1 x AAX78245 ..

Align seg 1/1 to: AAX78245 from: 1 to: 4988

55 AsnteuProileaspLyaspaLaalysPboSerphealaaspSergingl 71
+++++ +++++ |||::||| ++++++
2131 GATGTTCTTATTACAGACGGAAACAACTWCACTGTGACTGCTTCA.. 2178
71 yelnTrpyrThrPropheilelaialaValaGlulysAlaglyalllel 88
||||||| |||::||| |||::|||
2179ACTGCATTTCGAATAAATTAATACGCAATTAAGAAGTAACTAGTTA 2218
88 ysGIYThGlyAsnGlyPheGluProAsnglyLySIleasParValser 104
||||| :: ||||||| :++
2219 AAGGATTTAAA.....GACAAAAATGGCAAAGAATTTAAAGAAGCAT 2259
105 MetalaSerLeuleuValaGlualalyTyrlsyleuaspThrlysyValasnG 121
+++++ :+++++ :: |||::||| |
2260 GCATTTCGATTCGAAATCGAAATGATCGTACTTGCTCAAGAGCTTTGG 2309

121 yThrProAlaThr..... 125
 |||||
 2310 AACTAATGTAACAACAACACTTCTGTAACCTTAGCAGCAGGTACTTTCG 2359
 126 LysPheLysAspLeuGlu 131
 2360 ACACGACGACTACTTTACAGTAGATTTGATTAAGTTGTTACACCCTGAA 2409
 132 ThrLeuAsn..... TrpGlyLys.. 137
 |||||
 2410 ACTGTAAACAGCTCGAACCTTACTATTACAGATGTTGAACCTGGAAACG 2459
 138 GluLysAlaAsnIleLeuValGluLeuG 147
 2460 CATTCAGTAATGTCATCTACTTCTGCTTCTACAAATTACTATTAGCTTAA 2509
 147 LysIleSerValGlyThrGlyAspGlnTrpGlu..... ProlLys 159
 2510 AAGAAGCGTTAGTAACTGGTAAACAATATAAAGCTGTAATCAATATGCTT 2559
 160 LysThrValThr..... LysAlaGluAlaIleGlnPheIleAlaLysThr 174
 |||||
 2560 AAAAGCTTAAGCTGTTACAAATCAGAGAGCTTACGAGTTAGTG..... 2601
 174 rAspLysGlnPheGlyThrGluAlaIleLysValGluSerAlaLysAlaV 191
 2602 TTCACGCAAAACGACATACGACCAACTGTTGCTACCGCTC 2641
 191 aLThrThr..... 193
 2642 CTACTACTTCTAGCTGTACACTTATCTACTGTTCTTCTTACACAAT 2691
 193 193
 2692 GTTTGGCGTAATTTGGCTGTGTGTGATGAAGTGAAGCTGAATATTATTC 2741
 194 GluLysValGluValLysPheSerLysAlaValGluLysLeuThrL 209
 2742 TCGCTCTCAATTCACAAACACGTTGCTACTAAGTTAGACGAATCACTT 2791
 209 yGluAspIleLysValThrAsnLysAlaAsnAsnAspLysValLeuVal 225
 2792 TAGCGATTAACCTTGTATTAGTTGAAAAAGAAATGTGTACAGTTGTGCT 2841
 226 LysGluValThrLeuSerGluAspLysArgSerAlaThrValGluLeuTy 242
 2842 TCTGAACCTAAATATATATGCACAGCCTAAATGCTAATCTTAGTCCAAA 2891
 242 rSerAsnLeuAlaIleLysGlnThrTyThrValAspValAsnLysValG 259
 2892 ACCGGACCTTAAAGAAATATACAAATCTCAAAATTAAAGGCT 2941
 259 LysThrThrGlu..... ValAlaValGlySerLeuGluAlaLysThrIle 273
 2942 TGAAGTCGATTAACGATTTGAATTAGGACCTGTTAACGAGAAACATAT 2991
 274 GluMetAlaAspGlnThrValValAlaAspGluProThrAlaLeuGlnP 290
 2992 GAGTTCAAAACACAGACTTACTCTCT..... CCTACAGTTATTAGCCT 3035
 290 eThrValLysAsp..... GluAsnGlyThrGluValValSerProGluGlyI 306
 3036 AACGTCCTAAATAATGGCAGCGCTGATTTAAAGTAACT..... GAAGCTC 3079
 306 LeuPheValThrProAlaIleGluLysIleAsnAlaLysGlyLeuLeu 322
 3080 AAGAATTTACTGTGAAGTTCTCAGAGAAATTTAAATCAATTTAATGCTCA 3129
 323 ThrLeuAlaLysGlyThrSerThr..... 330
 3130 ACCGTTTCGGGTAGCAACATCAATACGCTCAAGTTGCTGTAGTAAAGC 3179

331 ThrV 332
 3180 GGGTCAAACTTATTCGCTTACAGACAAGTACATCATTCACAGCTAGTG 3229
 332 aLysAlaValTyrrLysAspGly..... LysValValAla 344
 3230 TTGAACGGTACTGCTGCAAGATGGAACATCAAGTGAAGTACTGCTCT 3279
 345 GluSerLysGlu..... 348
 3280 AACCAATTAGAAGCGTAACCAAGGTACAAATTAGTAGTTCGTGAAGG 3329
 349 ValLysValSerAlaGluGlyAlaIleValAlaSerI 361
 3330 TGCACACGCTCCTGTTAAAGATGCTGCAAAATGCAATCTTAGCAACTA 3379
 361 LeSerAsnTrpThrValAlaGluGlnAsnLysAlaAspPheThrSerLys 377
 3380 ACTATATCTATACATTTACAACTGAAGGTCAA... GACGTACACAGCCA 3426
 378 AspPheLysGlnAsnAsnLysValTyrrGluLysAspAsnAlaTyrrValG 394
 3427 ACGGTT..... ACAAAAGTATTCAAAGGTATTCCT..... 3456
 394 nValGluLeuLysAspGlnPheAsnAlaValThrThrGlyLysValGluT 411
 3457 TTAAAGACGCT... GATGCAGTTACTACA..... 3483
 411 yrrGluSerLeuAsnThrGluValAlaValAlaLysAlaLThrGlyLys 427
 3483 3483
 428 ValThrValLeuSerAlaGlyLysAlaProValLysValThrVal..... 442
 3484 CTTACGAACGTTGATGACAGGTCAA..... AATTCACATCACTAAT 3524
 443 LysAspSerLysGlyLysAlaLeuValSerHisThrV 455
 3525 TAGCGAAGAAATTAACAAACGATGAGTGT... TCTTATGAGGTGCAAG 3571
 455 aGluIleGluAlaPheAlaGlnLysAlaMetLysAspIleLysLeuGlu 471
 3572 TAACGTCTGAGAAATTAACAAACGATGAGTGTAT..... GCT 3612
 472 LysThrAsnValAlaLeuSerThrLysAspValThrAspLeuLysValLy 488
 3613 GSTACTGGAACACACTGATCAAGT..... 3636
 488 sAlaProValLeuAspGlnTyrrGlyLysGluPheThrAlaProValThrV 505
 3637 .GCTCCTAAGACAGATGCAAAATGTTAAAGTAAACGCTGCTGTGTACAT 3685
 505 aLysValLeuAspLysAspGlyLysGluLeuLysGluGlnLysLeuGlu 521
 3686 TAACGTGCTTGTACATTAACGACAAAGATGCGAAA... TTGCGTCTGTA 3732
 522 AlaLysTyrrValAsnArgLuleuValLeuAsnAlaIleGln..... 536
 3733 GTAGTAAAGCTCTTCTACATGATGGAATTCGTATGAGCTGTAATGTAAT 3782
 537 GluAlaGlyAsnTyrrThrValV 544
 3783 TAAGAAAAAGATATTTAAATTCGTTACAAACAGCTGAGCACACTGTAG 3832
 544 aLLeuThrAlaLysSerGlyLysGluAlaLysAlaThrLeuAlaLeu 560
 3833 CTTCGTGAAAGCTGCTGCTGACAAAGATGCTCAA..... 3867
 561 GluLeuLysAlaProGlyAlaPheSerLysPheGluValArgGlyLeuAs 577
 3868 AACGTTCTGCTGCAATCCCA..... ACAG 3893
 577 pThrGluLeuAsp..... LysTyrrValThrGluGluAsnGlnLysA 591

```

3994 CACGCAATTGATCACTAGACGCTATTAGTGAATTCATGAA... 3939
591 snAlaMetThrValSerValLeuProValAspAlaAsnGlyLeuValLeu 607
3940 .....ACTGATTACGGGAAGTTAAACCTGACAGACATCGTTGTT 3978
608 LysGlyAlaGluAlaAlaGluLeuLysValThrThrAsnLysGluG 624
3979 AAAGATGACAGAGGTAATGCGTACTGCTGTAACAGATGACAGG 4028
624 yLysGluValAspAlaThrAspAlaGluValThrValAlaAsnAsnSerV 641
4029 TTCT.....ACAATAAATTG 4045
641 AlIleThrValGlyGlnGlyAlaLysAlaGlyGluThrTyrLysValThr 657
4046 TATTCACCTCCATCTCAAGATTAAAGCTGCTAGCTTACTCTTAACA 4095
658 ValValLeuAspGly.....LysLe 664
4096 ATT.....GACGTCGACAGATTAAGTAGTACACACATCTTAATA 4139
664 uIleThrThrIleSerPheLysValValAspThrAlaProThrAlaLysG 681
4140 CATTACT.....TCGTTCAAGACTGATCTGCCAATCCACG..... 4176
681 yLeuAlaValAlaGluPheSerThrSerLeuLysGluValAlaAlaProAsn 697
4177 .....TTATCTTCATCAGCATTTGCGACGGTGACATTAC 4212
698 AlaAspLeuLysAlaAlaLeuLeuAsnIleLeuSerValAspGlyValPr 714
4213 GTTGAC...CGTTCTAAACAATTACATTCAGATTCAGGATTCAGTTCC 4259
714 o...AlaThrThrAlaLysAlaThrAlaSerAsnValGluPheValSera 730
4260 AAACCCAAACAATCACTTTAAGAGGCTGACGGAATTCATTACTTAAT 4309
730 laAsp...ThrAsnValValAlaGluAsnGlyThr.....ValGly 742
4310 ACACCTTGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATC 4359
743 AlaLysGlyAlaThr.....SerIleTyrValLysAsnLeuThrValVa 757
4360 CACAAGGCTGAACACTTGACGAGCTTACTCAATATGAGTTAGCAGTTTC 4409
757 lLysAspGlyLys.....GluGlnLysValGluPheAspL 769
4410 AAAAGATTTCAAACTGCTACTGATATTGATGACAAAGTTACATTTC.... 4455
769 ysaAlaValAlaValAlaValSerIleLysGluAlaLysProAla 783
4456 .....ATCACAGGTTCTGTTGCTACTGACGAAGTAAACCTCTCT 4494
seq_name: /STDS2/9c9data/geneseq/NA1992.DAT:AAQ29471
seq_documentation_block:
ID AAQ29471 standard; DNA; 6744 BP.
XX
AC AAQ29471:
XX
XX 04-MAR-1993 (first entry)
DE Extracellular factor related protein gene.
XX
XX EF*: detection; prevention; screening; diagnostic; ss.
XX
XX Streptococcus suis type II (non-pathogenic).
XX
XX Key location/Qualifiers
XX CDS 361..5826
XX /tag- a
XX

```

```

FT sig_peptide 361..498
FT /tag- b
FT mat_peptide 499..5826
FT /tag- c
FT promoter 66..71
FT /tag- d
FT /note- "-35 region"
FT 89..94
FT /tag- e
FT /note- "-10 region"
FT 153..158
FT /tag- f
FT /note- "-35 region"
FT 176..181
FT /tag- g
FT /note- "-10 region"
FT 350..356
FT /tag- h
FT /note- "ribosome binding site"
FT 2869
FT /tag- i
FT /note- "start of repetitive unit R1"
FT 3097
FT /tag- j
FT /note- "start of repetitive unit R2"
FT 3292
FT /tag- k
FT /note- "start of repetitive unit R3"
FT 3520
FT /tag- l
FT /note- "start of repetitive unit R4"
FT 4087
FT /tag- m
FT /note- "start of repetitive unit R5"
FT 4381
FT /tag- n
FT /note- "start of repetitive unit R6"
FT 4609
FT /tag- o
FT /note- "start of repetitive unit R7"
FT 4837
FT /tag- p
FT /note- "start of repetitive unit R8"
FT 5065
FT /tag- q
FT /note- "start of repetitive unit R9"
FT 5293
FT /tag- r
FT /note- "start of repetitive unit R10"
FT 5521
FT /tag- s
FT /note- "start of repetitive unit R11"
FT 6554..6566
FT /tag- t
FT /note- "dyad symmetry region 1a"
FT 6571..6583
FT /tag- u
FT /note- "dyad symmetry region 1b"
FT 6611..6625
FT /tag- v
FT /note- "dyad symmetry region 2a"
FT 6631..6644
FT /tag- w
FT /note- "dyad symmetry region 2b"
FT
XX
XX W09216630-A.
XX
XX 01-OCT-1992.
XX
XX 19-MAR-1992;
XX
XX 21-MAR-1991;
XX
92WO-NL00054.
91NL-0000510.

```

136 CLYGLYSCULYSLAASNIIELEUVALGLULEUGLYLESERVALGLY.. 151

3520 CAGAAAGATGCTAGACAACAAGTTTCCCAAAAGACGAAGTACAGCTAAAGA 3565
152ThrGlyAspGlnTrpGluProLysL 160
3570 AGCGATTTGACGCAACATCCAACTTACCTCAGATGAGAAAGAAATCAGCTA 3619
160 yStThrValThrIlysValGluValAlaGlnPheIleAlaLysThrAspLys 176
3620 ACAAATGCACTGGAAAGAGCGCGCTTAAGGTAGACACACGCCGCTATTGTATAA 3665
177 GlnPheGlyThrGluValAlaLysValGluSerAlaValAlaValThr.. 192
3670 GCATCAACTCAGATGACAGTTCACTCAATAGAAAGAGCAAAAGGTAGACGC 3719
193ThrGlnLysValGluValLys..... 199
3720 TATCAATTTGATTACTGTCGCAAGCGAAGTGTAAAGTGTCACTTCTGCTGA 3765
200PheSer 201
3770 ACTTGGCAGATCAATCAAGAAGCTCGAAGATTAACCAAGCAAGAAAGCAAGA 3819
202 LysAlaValAlaGlu.....LysLeuThrLysGluAspIleLysValThrAs 216
3820 AAAGCTATCGAATGCGCTCAACTATGCTACTATGCGGAAGCAATCGCTTAA 3869
216 nLysLysAlaAsnAspLysValLeuValLysGluValThrLeuSerGluA 233
3870 GAAGGCTCTTCAAGATGTTGTGTGATAAAGAAAGAACACAGCTTGACAGCG 3919
233 sPLysArgSerAlaThrValGluLeuTyr..... 242
3920 CAGCTAAGGGTACCAACAATGATTCATGAAAGCTATTACTACAGAAAAA 3969
243SerAsnLeuAlaAlaLysGlnThrTyr..... 252
3970 GCGAAAGCGCGCGAAGCTTGTCTGGCAAAAGACTTGCACAGACACAGTTAA 4019
253ValAspValAsnLys.....ValGlyLysT 261
4020 AGAAGCTGAGATGACGATTTGAATTTGGCTAAAGATTAAGATTAAGTAAAG 4069
261 hGluValAlaValGlySerLeuGluAlaLysThrIle.....GluMet 275
4070 AAGCAATCCGAACAGAAAGAGAGCTACTAAATTAAGTAGAGAAACTT 4119
276 AlaAspGlnThr.....ValValAlaAspGluProThrAlaLeuGlnPh 290
4120 CGAGAAGATAGCGCGAAGAGCTATCGAGACAAATCA..... 4155
290 eThrValLysAspGluAsnGly.....ThrGluVal 300
4156 AACTGTGCAGATGAAGATTAAGCAGCGGAAATTAAGAAAGCTAACTGCG 4204
300 alValSerPro.....GluGlyIleGluPheValTyr 310
4205 CHTGGCAAAAACCTTACCAACATGCGTGACAACTGCAGATTAACGTAACG 4254
311 ProAlaIaGluLysIleAsnAlaLysGlyLuiIleThrLeuAlaLysG 327
4255 CAAGAGACGAAAGAAAGCTCAACCCCTTACCAATCTTGAAGAAAGCTAAAGA 4304
327 yThrSerThrThrValLys.....AlaValTyrL 337
4305 AACACAGAAATTTGCAGATTAAGCTCGCATTTGATAGTTGACTTACTTGG 4354
337 ySLysAspGlyLysValAlaIaGluSerLysGluValLys..... 350
4355 TGAAGATGTCGACGCTTAACGCTACTAAACAAGATCTGAACACCAAGATT 4404
351ValSerAlaIaGluGlyValAlaValAlaSerIleSerAsnTr 364
4405 GCTAAAGATGACGCGCTGCTTAAGAAAGAACCAATTTCAAGCATCCAACTT 4454

```

364 pThvAlaGluGlnAsnLysAlaAspPheThr.....SerL 377
    ||| ||||| ||| |||||
445 GACGACGCGAG.....AAGAAACCTTCACGATGCGTAGATGCAG 4498
377 ysAspPheLysGlnAsn.....Lys 384
    ||| ||| ||| |||
4499 AAGTAGCCAAAGCTAACGACCAATTTCAGTCGACACCGCCAGCAT 4548
385 ValTyrGluGlyAspAsnAlaTyrValGlnValGluLeuLysAspGlnPh 401
    ||| :::: |||| ||| :::: |||||
4549 GTTCAAAAAGAGAGAGATGCGAGTGTTCACAGCATTCGAGAAGATGTTCT 4598
401 eAsnAlaVal.....ThrThrGlyLysValGluTyrGluSerLeuA 415
    :::: ||||| :::: |||||
4599 TGACGCGAGCTAAACAGATGCTAAGAAATTAAGATTGCTAAAGATGCAGCG 4648
415 snThrGluValAlaVal..... 420
    :::: |||||
4649 CTGCTAAAGAACAAATTGGCTCCAAATCCAAACTGCAGACGACGAGAG 4698
421 .....ValAspLysAlaThrGlyLysValTyr...Va 430
    ||||| |||||
4699 AAACCTTACCGATGGCGGTAGATGACAGAGTACCAAGCTAACGACGC 4748
430 lLeuSerAlaGlyLysAlaProValLysValThrValLysAspSerLysG 447
    :::: ||||| :::: |||||
4749 AATTTCAGCTGCCAACCAAGCCAGCAGATGTTCAAAAAGAGAGATGTCAG 4798
447 lLysAlaLeuValSerHisThrValGluIleGluAlaPheAlaGlnLys 463
    || ||| :::: |||||
4799 GTGTTCGACGCAATTCAGAGAGATTT...CTTGACGCGAGCTAAACAGAT 4845
464 AlameLysAspIleLysLeuGluLysThrAsnValAlaLeuSerThrLy 480
    ||| :::: |||| ||| :::: |||||
4846 GCTAAGATAGATTGCTTAAAGAA...TTCGACGCTGCTAAGTCCAGCAT 4892
480 sAspValThrAspLeuLysValLysAlaProValLeuAspGlnTyrGlyL 497
    ||||| :::: |||||
4893 TGACGCGCAT.....CCAACTTGACAGATGCGAGAG 4924
497 ysgLupPheThr.....AlaProValThrValLys 506
    ||||| :::: |||||
4925 AGGAATCGCTAAGAACACAGCTGATGCTGAATGAAGTGCAGCAGAT 4974
507 ValLeuAspLysAspGlyLysGluLeuLysGluGlnLysLeuAlaLys 523
    :::: ||||| :::: |||||
4975 GCAATTGATGCTTCAACAAGTCCAGTCAAGCGCATCGGACAGAGACAA 5024
523 sTyrValAsn.....ArgGluLeuValLeuAsnAlaAlaGlyGlnGluA 538
    ||| ||||| :::: |||||
5025 AGGCGTAGGTTCAATCCGCCCAAGATGTTCTTGACGCGACCAACAGATG 5074
538 lAglYAsnTyrThrValValLeuThrAlaLysSerGlyGlnLysLysLys 554
    || ||| |||||
5075 CTAAGAAC.....AAGATTGCCAAAGAAAGTTGCC 5103
555 LysAlaThrLeuAlaLeuGluLeuLysAlaProGlyAlaPheSerLysPh 571
    ||| ||||| :::: |||||
5104 GCAGCTAAGAACGATGATGCCAATCCG..... 5133
571 eGluValAlaGlyLeuAspThrGluLeuAspLysTyrValThrGluGluA 588
    :::: ||||| :::: |||||
5134 .AACTTATCATGATGACAGAAAGAGCTTCTAAGAAACGCGTAGATGCAG 5182
588 snGlnLys.....AsnAlaMetThrValSerValLeuProValAsp 601
    ||| ||||| :::: |||||
5183 ATGCTAAGCTACGACACAGATGCAATGATGCTCAACAAGTCCAGTCCAA 5232
602 AlaAsnGlyLeuValLeuLysGlyValA..... 610
    ||||| :::: |||||
5233 GCGCAATGCGAGAGACAAAGCGCTAGTTCATCGCCCAAGATGTTCT 5282

```

```

611 .GluAlaAlaGluLeuLysValThrThrThrAsnLysGluGlyLysGluV 627
    :::: ||||| ||| :::: |||||
5283 TGACGCGAGCG.....AAACAGATGCTAAGCAATTAAGATTGCTAAAGAT 5326
627 alaAspAlaThrAspAlaGlnValThrValGlnAsnAsnSerValIleThr 643
    ||||| :::: |||||
5327 CCGACGCTGCTAAGTCAAGCATTGACCGCAATCCAACTTGACAGATGCA 5376
644 ValGlyGlnGlyAlaLysAlaGlyGluThrTyrLysValThrValLe 660
    :::: |||||
5377 GAGAAGCAATCAGCTAAG..... 5394
660 uAspGlyLysLeuIleThrThrHisSerPheLysValValAspThrAlaP 677
    ||||| :::: |||||
5395 .....AAGCGGTAGAT..... 5406
677 roThrAlaLysGlyLeuAlaValGluPheThrSerThrSerLeuLysGlu 693
5406 ..... 5406
694 ValAlaProAsnAlaAspLeuLysAlaAlaLeuLeuAsnIleLeuSerVa 710
    ||||| |||||
5407 .....GCAGATGCTAAAGCTGCG.....AC 5426
710 lAspGlyValProAlaThrThrAlaLysAlaThrAlaSerAsnValGlu. 726
    ||||| :::: |||||
5427 AGATGCAATTGATGCTTCAACAAGTCCAGTCGAAAGCCCAATCGCAGAGG 5476
727 .....PheValSerAlaAsp..... 731
5477 ACAAGGCGTAGGCGCCATCGCCAAAGACATTTGATGCCGCCAAACAA 5526
732 .....ThrAsnValValAlaGluAsnGlyThrValGluAlaLysGlyAl 746
    ||| :::: |||||
5527 GATGCTAAGAACAGATGCTTAAAGAG.....CGACGATCCGC 5564
746 aThrSerIleTyrValLysAsnLeuThrValValLysAspGlyLysGlu. 762
    ||||| :::: |||||
5565 TAACTCAGTCTAATGACCTCAATCCGAACCTTGACAGATGCACTAAGAG 5614
763 .GlnLysValGluPheAspLysAlaValGlnValAlaVal 775
    ||| |||||
5615 CGGCTAAATCTGAATATGATTAAGCTGTTGAGAAAGCATT 5655
seq_name: /SID2/gcgdata/geneseq/geneseqn/NA1996.DAT.AAT41476
seq_documentation_block:
ID AAT41476 standard; DNM; 7291 BP.
XX
AC AAT41476;
XX
DT 15-JAN-1997 (first entry)
XX
DE Haemophilus adhesion protein HA2 gene.
XX
KW Haemophilus adhesion protein; HA2; hsf protein; vaccine; ss.
OS Haemophilus influenzae type b strain C54.
XX
FH key Location/Qualifiers
FT misc_difference 4
FT /*tag= a
FT /note= "base 4 is given as n in the specification"
FT misc_difference 57
FT /*tag= b
FT /note= "base 57 is given as n in the specification"
FT misc_difference 59
FT /*tag= c
FT /note= "base 59 is given as n in the specification"
FT CDS 163..7224
FT FT
XX /*tag= d
XX
PN W09630519-AL.

```


CC	2255	AAACGAGACGACT.....GATCGAGATCGCGGTAAAGTAAC	2292
CC	691	LeuLVaGluValAlaProasn...AlaAspLeuLVaAlaAlaLeuLeuAs	706
CC	2293	GTAAAAAGATCTACTCTCTAATGACGCTGATAGAAAGTCGCAACTGTAA	2342
CC	706	nileLeuSerValAspGlyValProAlaThrThrAlaLVaAlaThrAlas	723
CC	2343	AGATGT.....GCAACCGCAATTAATAGTCGCGCA	2374
CC	723	erAsnValGluPheValSerAlaAspThrAsnValValAlaGluVaangly	739
CC	2375	CTTTTGTGAAACAGAGAAATTAACCTACTCTATGATGAAGTAAATCCT	2424
CC	740	ThrValGluValAlas.....GlyAlaThrSerIleLeu	750
CC	2425	ACGAGTAACGGCAGAAAGATGACGACTTAAAGCGGCGATACCTTAACCT	2474
CC	750	rVal.....LysAsnLeuThrValValAlaLVaLysAspGlyLysGluGlnLys	765
CC	2475	TAAAGCAGGTAAAAACCTGAAGTTAAACGTGATGAAAA.....ATA	2518
CC	765	aGluPheAspLysAlaValGlnValAlaValSerIleLysGluAlaLys	781
CC	2519	TTACTTTTGACTTGGCGCAAAAC.....CTTGAGGTGAACCTGCGAAA	2562
CC	seq_name: /SIDS2/gcdata/geneseq/geneseqn/NA2000.DAT:AAA92489		
seq_documentation_block:			
ID	AAA92499	standard; DNA; 7253 BP.	
AC	AAA92499;		
XX	17-JAN-2001	(first entry)	
XX	Haemophilus influenzae adhesin (Hia) gene from type c strain API.		
XX	Hia: adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;		
XX	non-typable Haemophilus influenzae; antiinflammatory; auditory;		
XX	antibacterial; meningitis; epiglottitis; septicaemia; otitis media;		
XX	diagnosis; immunogenic; antigen; ds.		
XX	Haemophilus influenzae.		
XX	WO200055191-A2.		
XX	21-SEP-2000.		
XX	16-MAR-2000; 2000WO-CA00289.		
XX	16-MAR-1999; 99US-0268347.		
XX	(CONN-) CONNAUGHT LAB LTD.		
XX	Loosmore SM, Yang Y, Klein MH;		
XX	WPI: 2000-618897/59.		
XX	P-PSDB; AAB23860.		
XX	Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for		
XX	use as antigens and vaccines and for treating Hemophilus influenzae		
XX	infection		
XX	Claim 1; Fig 24; 275bp; English.		
XX	The present sequence represents a Haemophilus influenzae adhesin		
XX	(Hia) gene from the type c Haemophilus influenzae strain API.		
XX	Hia genes and proteins have antiinflammatory, auditory and antibacteri		
XX	activities, and can be used in the production of a vaccine. An		
XX	immunogenic composition comprising an Hia gene, a polypeptide encoded		
XX	by an Hia gene, or a recombinant Hia polypeptide is useful for inducing		
XX	protection against disease caused by Haemophilus strains in a		

CC susceptible host, preferably a human. An Hia protein is useful as an
 CC antigen, in immunogenic preparations including vaccines, as a carrier
 CC for other immunogens, and in the generation of diagnostic reagents. Hia
 CC is useful for treating diseases caused by the infection of Haemophilus
 CC influenzae such as meningitis, epiglottitis, septicemia and otitis
 CC media. Recombinant production of Hia favours high recovery of the
 CC protein compared to the recovery of native protein from Haemophilus
 CC influenzae species. A truncated protein has a significantly higher
 CC amount of recovery than a full-length protein.
 CC
 XX

Sequence 7253 BP: 2495 A; 1338 C; 1782 G; 1638 T; 0 other;

alignment_scores:

Quality: 238.50 Length: 849
 Ratio: 0.575 Gaps: 42
 Percent Similarity: 48.881 Percent Identity: 22.968

alignment_block:

US-09-754-947-1 x AAA92499 ..

Align seg 1/1 to: AAA92499 from: 1 to: 7253

23 GluLysGlyAlaValLysGlyAsnAspLysGlyMetPheGluProGlyLys 39
 |||||
 235 GAAAAAGAAAGTTACAGAAATTCAAATGGGAAATATATTCCACATATA 284
 39 sGluLeuThrArgAlaGluAlaAlaThrMetLeuAlaGlnIleLeuAsnL 56
 | : : : : : ||| : : : : : ||| : : : : :
 285 AGGAGTACTAAAGAACCGGAGCAATCCCTCAAGCGCGCAC..AAC 331
 56 euProIleAspLysAspAlaLys...ProSerPheAlaAspSerGlnGly 71
 || ||| : : : : : : : ||| : : : : :
 332 TGAAAAATCAACAAAGCACCAATGCCAGTACCTCACTACTCGTGAAA 381
 72 GlnTPtyrThrProPheIleAla...AlaValGluLysAlaGlyVal... 86
 : : : ||| : : : : : : : : : : : : : : : :
 382 AAAGACCTCAGATCTGACAGTGTGCACTGAAATAATTCGTTGG 431
 87
 432 CGCAAAACGCGATAAAGTTGATATTACCGATGCAATGCGCTGAAT 481
 96ProAsnGlyLysIle..... 100
 482 TGGCGAAAACAGATACGGAATGTTCATTTGAATGCTTGGATTCAC 531
 101AspArgValSerMetAlaSerLeuLeuValGluAlaTyrLysLe 115
 ||| ||| : : : : : : : : : : : : : : : :
 532 TTGCTGATGCGGTACGATACGATGAGTGTAAAGTTCAAGTTTAC 581
 115 uAspThrLysValAsnGlyThrProAlaThrLysPheLysAspLeuGluT 132
 : : : : : ||| : : : : : ||| : : : : :
 582 ACCTAATGATGTTGAAAAAACAAGAGCTGCACACTGTTAAAGATGTTTA 631
 132 hLeuAsnTyrGlyLysGluLysAlaAsnIleLeuValGluLeuGlyLe 148
 : : : : : : : : : : : : : : : : :
 632 ATGACGCTTGGAACATTTAAAGTGCTAAAACT..... 663
 149 SerValGlyThrGlyAspGlnThrProLysLysThrValThrLysAl 165
 : : : : : : : : : : : : : : : : :
 664GCTGGAGGTAAATGTTGAGAGTGTGATTAGTTCGCGTTA 704
 165 agLualAlaGlnPheIleAlaLysThrAspLysGlnPheGlyThrGluA 182
 : : : : : ||| : : : : : : : : : : :
 705 TAAATATGTTGAATTTATTACAGCGCATAAACACGCTTGATGTTGAT 754
 182 lAlaLysValGluSerAlaLysAlaValThrThrGlnLysValGluVal 198
 : : : : : ||| : : : : : : : : : : :
 755 TAAACAGCTAAAGAAAACGTTAAACA.....ACGCAAGTG 789
 199 LysPheSerLysAlaValGluLysLeuThrLysGluAspIleLysVal... 214
 ||||| : : : : : : : : : : : ||| : : : : :
 1576

790 AATTTCACACCGAAAAACCTCTGTATCAAAAGAAAAAGCGTAACTATT 839
 215 ThrAsnLysAlaAsnAsp.....LysValLeuValLysGluValT 229
 ||||| : : : : : ||| : : : : :
 840 TCTGGAAGAGAGATTAACGACAAATTAAGTTTCACTACAGCGGCA 889
 229 hLeuSerGluAspLysArgSerAlaThrValGluLeuTyrSerAsnLeu 245
 || : : : : : ||| : : : : : ||| : : : : :
 890 CTGATATATCAAGATGAGGATATGGCTTAGCTACGCAAAAGCTGTGAT 939
 246 AlaAlaLysGlnThrTyrThrValAspValAsnLysValGly..... 259
 ||| : : : : : : : : : : :
 940 GATGCT.....GTGAACAAGCTGTGAGAGAGT 968
 260 LysThrGluValAla.....ValGlyLys 267
 ||||| : : : : : ||| : : : : :
 969 TAAACCACTACTGCTTAATGTCAAATGGCAGCTTCCCACTGTGGCT 1018
 267 erLeuGluAlaLysThrIleGluMetAlaAspGlnThrValAlaAsp 283
 || : : : : : ||| : : : : : ||| : : : : :
 1019 CAGCACAATGTAACTTTGAAAGTGGGATGTACAAACAGCTCA... 1065
 284 GluProThrAlaLeuGlnPheThrValLysAspGluAsnGlyThrGluVal 300
 1066GTAACTAAAGATTAACGCAATGCGAT 1094
 300 lValSerPro.....GluGlyIleGluPheValThrP 311
 : : : : : : : : : : : : : : : : :
 1095 CACTGTAAAGTACGACGCGAAGTTGGGACGCGCTGAATTT.....G 1138
 311 roAlaAlaGluLysIleAsnAlaLysGlyLuiIleThrLeuAlaLysGly 327
 : : : : : ||| : : : : : : : : : : :
 1139 ATAGCGATAAAAAATC.....GTTGCAGATACG 1167
 328 ThrSerThrThrValLysAlaValTyrLysLysAspGlyLysValValAl 344
 ||||| : : : : : ||| : : : : :
 1168 ACCGCCTTACTGTGACAGT.....GTAAGGTAGCGGA 1202
 344 agLysSerLysGlu.....ValLysValSerAlaGluGluAlaVal 358
 : : : : : ||| : : : : : ||| : : : : :
 1203 AATGCTTAAGAGATGACAAAGAAAACTGTTAATGACGCGATTGG 1252
 358 AlaIleSerIleSerAsnTyrThrValAlaGluGlnAsnLysAlaAspHe 374
 ||||| : : : : : ||| : : : : : ||| : : : : :
 1253 TAACAGCTTTAGTAACTAAAGTGAAGCAAAAGCTGAGGCTCAT... 1299
 375 ThrSerLysAsp.....PheLysGlnAsnAsnLysValTyr 386
 ||||| : : : : : : : : : : : : : : : :
 1300 ACTGATCTGATGTCGCGCTGACGGGATTTCAAAAGCAAGAGTCAA 1349
 386 rGluGlyAspAsnAlaTyrValGlnValGluLeuLysAspGlnPheAsnA 403
 ||||| : : : : : ||| : : : : : ||| : : : : :
 1350 AGCAGCGCAAAACGTA.....ACCTTAAAG 1375
 403 lAlaValThrGlyLysValGluTyrGluSerLeuAsnThrGluValAla 419
 : : : : : ||| : : : : : : : : : : :
 1376 CGGCGAAGACATTAAAGTGAAGACAGATGGCGCACTTAATTATCA 1425
 420 ValValAspLysAlaThrGlyLysValThrVal.....Le 431
 : : : : : ||| : : : : : : : : : : :
 1426 CTGCAAGATGCTTTAAGCGGTTTAAAGACATTAAGTGGTGAAC 1475
 431 uSerAlaGlyLysAlaProValLysValThrValLysAsp..... 444
 : : : : : ||| : : : : : ||| : : : : :
 1476 TAAATGCGGAATATGATGGGAAACCGTATCAACAAACGTTTAAACA 1525
 445SerLysGlyLysAlaLeuValSerHisThrValGluIle 457
 : : : : : ||| : : : : : : : : : : :
 1526 TCACGCGACAGGATATGCGGTAGACAGATCAACAAACATGACGCTA 1575
 458 GluAlaPheAlaGlnLysAla...MetLysAspIleLysLeuGluLysTh 473
 : : : : : ||| : : : : : ||| : : : : :
 1576 ACCAAAGATGCAATTAAGCAGGTAATTAAGCTAAT.....AC 1613


```

473 rAsnValAlaLeuSerThrIysAspValThrAspLeuValLysAlaP 490
||||| : : : |||
1614 TAAATGTCGAGTGTAAAGACCTAATGACGAT..... 1647
490 roValLeuAspGlnTyrGlyLysGluPheThrAlaProValThrValLys 506
: : : : :
1648 .....GCGAATTTCAT 1659
507 ValLeuAspLysAspGlyLysGluLeuLysGluGln..... 518
||||| : : : : :
1660 GTTTAAATATACTCTGCACATGATTAAATAGACACGTTGAAGATGCTTA 1709
519 .....LysLeuGluAlaLysTyrValAsnArgGlu...LeuV 530
: : : : :
1710 TAAAGTTTATTAAATCTAAATGAAAAAATGCAAAATTAACACCGTTGG 1759
530 alLeuAsnAlaAlaGlyGlnGluAlaGlyAsnTyrThrVal..... 543
: : : : :
1760 TGACTGACAGCAGCGCGGCGACTGAGCGCATTTACGTAAATTGGTTGG 1809
544 ValLeuThrAlaLysSerGlyLysGluAla.....LysAlaThrLe 558
||||| : : : : :
1810 GTAGTATCAACCAAAACGATACGAAAGACAAACATCAAGTTAAACA 1859
558 uAlaLeuGluLeuLysAlaProGlyAlaPheSerLysPheGluValArg 575
: : : : :
1860 AGCTGATGAAGTCTTACCGGAGCC.....G 1888
575 LysLeuAspThrGluLeuAspLysTyrValThrGluGluAsnGlnLysAsn 591
||||| : : : : :
1889 GTGCTGCTACGCTTACTTCCAA.....TCGAAAAACGTTAAACAT 1929
592 AlameThrValSerValLeuProValAspAlaAsnGlyLeuValLeuLys 608
: : : : :
1930 ACGATTACCGTACTAGTGGCTGAACCTAAACGAGTAGCGGCTCTGAAA 1979
608 sGlyAlaGluAlaAlaGluLeuLysValThrThrAsnLysGluGlyL 625
: : : : :
1980 AGAGGCGGATCTATTAGCTCAAGTGATGATCAAAAC..... 2019
625 ysgLValAspAlaThrAspAlaGlnValThrValGlnAsnSerVal 641
||||| : : : : :
2020 .....ACTGATATATGTTTAACTGTGGTAATAAT..... 2049
642 IleThrValGlyGlnGlyAlaLysAlaGlyLysThrValThrVal 658
2049 ..... 2049
658 lValLeuAspGlyLysLeuIleThrThrHisSerPheLysValAlaSP 675
: : : : :
2050 .....GTTACTGCTGTCTACTAAAGGTGGCTTGAACCTGTAAA 2089
675 hAlaProThrAlaLysGlyLeuAlaValGluPheThrSerThrSerLeu 691
||||| : : : : :
2090 CTGAGAGGACT.....GATGCGATCCGCGGTAAAGTAACTGTA 2127
692 LysGluValAlaProAsn...AlaAspLeuLysAlaAlaLeuLeuAsnI 707
||||| : : : : :
2128 AAAGATGCTACTGTAATGACGCTGATAGAAGACCGCACTGTAAAGA 2177
707 eLeuSerValAspGlyValProAlaThrThrAlaLysAlaThrLase 724
: : : : :
2178 TGT.....GCCACGCAATTAATAGTCGGCGACTT 2209
724 snValGluPheValSerAlaAspThrAsnValAlaAlaGluAsnGlyThr 740
||||| : : : : :
2210 TTGTGAAAACAGACAATTAACTACTCTATTGTGAAGATAATCTTACA 2259
741 ValGlyValLys.....GlyAlaThrSerIleTyrVal 751
: : : : :
2260 GATTAACGGCAAAAGATGACGCACTTAAAGCGGCGATACCTTAACCTTAA 2309

```

```

751 l.....LysAsnLeuThrValValLysAspGlyLysGluGlnLysValG 766
||||| : : : : :
2310 AGCAGTAAAAAACCTGAAGCTTAAACGTGATGAGAAA.....AATATTA 2353
766 LuPheAspLysAlaValGlnValAlaValSerIleLysGluAlaLys 781
||||| : : : : :
2354 CTTTGACTTGGCGAAAAC.....CTTGAGGTGAATACTGCGAAA 2384

seq_name: /sids2/gcgdata/geneseq/geneseqn/NA1991.DAT:AAQ11789
seq_documentation block:
ID AAQ11789 standard; DNA; 4330 BP.
XX
AC AAQ11789;
XX
DT 01-AUG-1991 (first entry)
XX
DE HWP protein gene.
XX
KW HWP; protein; food; ss.
XX
OS Bacillus brevis HPD 31.
XX
FH key Location/Qualifiers
FT CDS 548..3895
FT /tag= a
FT /product= HWP
FT mat_peptide 707..3895
FT /tag= b
FT sig_peptide 548..707
FT /tag= c
FT sig_peptide 638..707
FT /tag= d
FT repeat_region 3937..3971
FT /tag= e
FT repeat_region 4174..4202
FT /tag= f
FT RBS 529..540
FT /tag= g
FT RBS 617..627
FT /tag= h
FT -35_signal 245..250
FT /tag= i
FT -10_signal 268..273
FT /tag= j
FT misc_signal 280..280
FT /tag= k
FT /function= transcription initiation site
FT /note= "p1"
FT -35_signal 359..364
FT /tag= l
FT -10_signal 382..387
FT /tag= m
FT misc_signal 392..392
FT /tag= n
FT /function= transcription initiation site
FT /note= "p2"
FT -35_signal 427..432
FT /tag= o
FT -10_signal 450..455
FT /tag= p
FT misc_signal 462..462
FT /tag= q

```


92x2